



Predicting Proportionality: The Case for Algorithmic Sentencing

Vincent Chiao

To cite this article: Vincent Chiao (2018) Predicting Proportionality: The Case for Algorithmic Sentencing, Criminal Justice Ethics, 37:3, 238-261, DOI: [10.1080/0731129X.2018.1552359](https://doi.org/10.1080/0731129X.2018.1552359)

To link to this article: <https://doi.org/10.1080/0731129X.2018.1552359>



Published online: 13 Dec 2018.



Submit your article to this journal [↗](#)



Article views: 752



View related articles [↗](#)



View Crossmark data [↗](#)

ARTICLE



Predicting Proportionality: The Case for Algorithmic Sentencing

VINCENT CHIAO*

A basic principle in sentencing offenders is proportionality. However, proportionality judgments are often left to the discretion of the judge, raising familiar concerns of arbitrariness and bias. This paper considers the case for systematizing judgments of proportionality in sentencing by means of an algorithm. The aim of such an algorithm would be to predict what a judge in that jurisdiction would regard as a proportionate sentence in a particular case. A predictive algorithm of this kind would not necessarily undermine justice in individual cases, is consistent with a particularistic account of moral judgment, and is attractive even in the face of uncertainty as to the legitimate purposes of punishment.

Keywords: proportionality, sentencing, judicial discretion, arbitrariness, algorithms

Sentencing is a bit of a black art. In imposing a just and appropriate

sentence, the sentencing judge must have in view a wide array of qualitatively distinct facts pertaining to the case before her—the nature of the crime, the harm inflicted, the suffering of the victim, the prospects for rehabilitation, society’s interest in condemning the crime, the offender’s prior criminal history, the offender’s familial obligations, her employment status, and so on. The sentencing judge must take these myriad facts into account and translate them into a concrete punishment, one that is neither disproportionately harsh nor overly lenient, and that is also in keeping with what other similarly situated offenders have received. How this

*Vincent Chiao is a Associate Professor, Faculty of Law, University of Toronto. Email: vincent.chiao@utoronto.ca

[I am grateful to audiences at the Centre for the Study of Law and Society at Berkeley Law School, King’s College London, Osgoode Hall Law School, and the Centre for Ethics at the University of Toronto for comments on earlier drafts. I received invaluable written feedback from Aziz Huq as well as from an anonymous referee for *Criminal Justice Ethics*. I am also indebted to Ioana Dragalin for editorial assistance.]

translation happens is rather mysterious. Offhand, many different punishments might seem appropriate in the abstract. That is what makes sentencing into a black art.

Perhaps unsurprisingly, in many jurisdictions—including Canada and much of the United States—sentencing remains by and large a highly discretionary affair. Setting aside the controversial issue of mandatory minimums, in many places sentencing remains a retail-level decision that is highly dependent upon the professional judgment of a number of people, but particularly that of the sentencing judge. While appellate courts review sentences, generally speaking they do so with a high level of deference to the original sentence. Hence, there is a high premium for experience. Experienced lawyers gradually develop a rough sense of the going rate for common types of cases in this or that courthouse, or even before this or that judge.

The highly discretionary character of sentencing raises obvious questions about fairness. It is far from obvious that like cases are being treated alike. Indeed, the system seems set up to ensure that they are not. Although extreme outliers are likely to be reversed on appeal, that still leaves substantial room for inter-court, inter-judge and even intra-judge variability. In the case of Canada, there is very little by way of systematic data about sentencing, so it is hard to be precise. For serious crimes carrying lengthy custodial sentences, however, it seems plausible to believe that the range of acceptable variation (i.e. before a sentence becomes likely to be reversed on appeal) can be a matter of many months—and, for the most serious

crimes, perhaps even years.¹ Hence, the liberty interests at stake are far from trivial.

Moreover, in addition to concerns about comparative fairness, concerns about arbitrariness are also frequently voiced in discussions about sentencing. By now, reports indicating that judges and juries exhibit predictable biases of various kinds are routine. From the racial or gender identity of the parties, to the time of day, attractiveness of the accused or victim, or (in the United States) proximity to an election, social scientists have documented a wide range of deeply troubling indicators of bias in sentencing. This is a distinct concern from comparative fairness. Suppose that judges uniformly set higher sentences in the afternoon than in the morning, and that everyone faces equal *ex ante* chances of being sentenced in the morning or in the afternoon. That kind of regime might be comparatively fair (each person is treated similarly), but it would still be substantively arbitrary, in the sense of introducing a bias that is not connected to culpability. After all, the reasons for the sentence would be unrelated to the crime.

This is not a merely theoretical concern. The incarcerated population in both the United States and Canada exhibits a high degree of racial skew, particularly in the case of African-Americans and Latinos (in the US) and indigenous people (in Canada.) There are lots of reasons for this disparity. One of them is that race plays, consciously or unconsciously, a role in determining who becomes a suspect, who is prosecuted, and how much punishment that person receives upon conviction. This suggests that our criminal

justice systems impose sentences in a substantially arbitrary manner. It also suggests that the sentences that are imposed are unfair. It is hard to imagine how we could adequately explain to a black defendant why he should accept a significantly harsher sentence than an otherwise identically situated white defendant. That just looks like discrimination.

None of this is surprising. People have been complaining about arbitrariness and bias in sentencing for a long time. I have been rehearsing some of these concerns because they are worth bearing in mind when thinking about possible alternatives to discretionary sentencing. It is easy to point out that no method for sentencing criminals—a difficult task under the best of circumstances—is perfect. But an alternative does not have to be perfect; it just has to be better than the status quo. And, when it comes to sentencing, the status quo does not inspire confidence.

What I shall be considering in this paper is a proposal to rely on recommendations generated by a machine-learning algorithm to supplement judicial discretion at sentencing. However, while support for sentencing algorithms is not unheard of the American Law Institute's Model Penal Code, for instance, has come out in favor of "actuarial-risk-assessment instruments" in

sentencing—my proposal is unusual in one important respect.² Most of the attention given to algorithmic decision-making in sentencing has focused on predicting whether a given defendant will re-offend in the future.³ In contrast, the function of the algorithm I consider in this paper is to predict what the modal judge in a given jurisdiction would regard as the proportionate sentence.⁴ In other words, it is designed to predict the behavior of *judges*, not that of *defendants*.⁵ Based upon a finite list of case-specific factors, the algorithm would generate a prediction of the sentence that would be regarded as proportionate for such a crime in the relevant jurisdiction, as well as sentences within a standard deviation from the average. Judges would not be bound by the recommendation, but would have a sense of what they and their peers have regarded as proportionate in other, similar, cases. Although there is little agreement about the purposes of punishment, in a sentencing regime that prioritizes proportionality, the algorithm's predictions would contribute to establishing a publicly-known and predictable baseline, one grounded in the judiciary's own sense of proportionality. I shall suggest that sentencing by the numbers may be better—less arbitrary and more fair—than sentencing without them.⁶

1.

Proportionality may take a weaker or a stronger guise. In its weaker form, proportionality states that an offender should not be punished more harshly than he or she deserves. In its stronger form, proportionality recommends

that people not be punished less harshly than they deserve, either. The distinction can be illustrated in terms of Norval Morris' well-known theory of limiting retributivism.⁷ According to that theory, desert sets

outer bounds of proportional punishment, with utilitarian or more pragmatic principles determining the specific punishment within those bounds. However, as Richard Frase has noted, while there appears to be substantial agreement that desert sets an upper bound on permissible punishment, it is more disputed whether proportionality constrains the minimum amount of punishment. As Frase points out, the American Law Institute (the drafters of the Model Penal Code) rejected imprisonment solely to vindicate proportionality; considerations of public safety would be required to justify any degree of imprisonment.⁸ This asymmetry between desert as setting an upper versus a lower limit is reflected in the distinction between weaker and stronger forms of proportionality.⁹ For the purposes of this paper, I will be concerned with the stronger form of proportionality, which sets both an upper and a lower bound on permissible punishment.

Proportionality is often believed to require a high degree of judicial discretion at sentencing. One reason is that a great many factors may be relevant to determining what kind of sentence would be proportionate in a given case. These factors range from the accused's state of mind at the time of the offence, his contrition (or lack thereof), the harm suffered by the victim, the seriousness of the type of crime at issue, the accused's prior record, the victim's role in instigating the crime, and so on. Any attempt to exhaustively specify these factors by statute would inevitably be under-inclusive.

Another reason that proportional sentencing is linked to discretionary sentencing is that even if we have identified all the relevant factors,

how they interact with each other can be difficult to specify in a general way. For instance, it may be that for a person to assault a family member is typically aggravating. Suppose, however, that a loving father ends the life of his severely disabled daughter in the belief that her life has been one of intense and constant pain with no hope of relief.¹⁰ His act may still be a crime, and it may still be punishable, but arguably in this context the fact that the accused is a family member is actually mitigating rather than aggravating; we understand it to be motivated by love and concern, rather than a desire to dominate and exploit. Hence, arguably one and the same factor—being closely related to the victim—is sometimes aggravating and sometimes mitigating. Distinguishing when it is one and when it is the other might be regarded as a hallmark of human moral judgment, something no algorithm could do.

Jonathan Dancy refers to this feature of moral judgment as "holism." As Dancy defines it, "holism" about moral reasons amounts to the view that "a feature that is a reason in one case may be no reason at all, or an opposite reason, in another."¹¹ Holism implies that assessing the total effect of the aggravating and mitigating factors in a given case will be a more complicated affair than simply summing up the weights assigned to each factor. A factor that cuts one way in one type of case may cut another way in a different type of case. Hence, it would be too crude to think that we can simply check off the presence or absence of factors in each given case. The factors that contribute to a proportionality judgment are often inter-related, since how they affect our

judgment depends upon the presence or absence of other factors. For instance, that the accused and victim are closely related is aggravating in most contexts, but not when it is conjoined with the kind of care and love that give us reason to be especially protective of family relationships.¹²

In light of holism, proportionality might be represented as a function taking any permutation of n factors as inputs and linking them to proportional sentences as outputs. Hence, proportionality is defined for $f_1 * f_2 * \dots * f_n$ possible permutations, for each distinct aggravating or mitigating factor, f , and with the value of f determined by how many possible values that factor can take. Each distinct permutation calls for a morally distinct judgment. Hence, if we have three such factors—harm caused, familial relation, and malevolent intent—then holism suggests that [harm, family, malevolent] be treated as morally distinct from [harm, family, not malevolent], not merely as a more aggravated version of the latter. This is what it means to say that the moral significance of a person's crime is assessed in light of all the relevant features of his crime taken together, rather than composed piecemeal out of the contribution of each of the individually relevant features.

To be sure, the idea that X is sometimes aggravating but sometimes not, conditional on the presence or absence of Y , does not show—as defenders of discretionary sentencing sometimes seem to suggest—that proportionality is ineffable, only that it is complex. Nevertheless, when taken together with the indefinite, but in any case, large, number of factors that might plausibly bear upon

proportionality, holism does suggest that sentencing is quite complex indeed. By way of illustration, suppose we have identified a dozen aggravating and mitigating factors, and that each of these factors could have one of four possible values. This would result in over 1.6 million distinct permutations of factors, each of which would have to be assessed separately. Even increasing the number of relevant factors modestly, to sixteen, yields over one billion possible permutations, and that is just for one type of crime. (What if someone is charged with more than one crime—how many possible combinations of charges would we need to consider? Is there a discount rate for multiple instances of the same crime? What if there are different victims? Etc.) Even if we might in principle create rules linking each permutation of factors to a proportionate sentence, the result would be so cumbersome that allowing judges to essentially intuit the correct sentence might seem reasonable in comparison.

Leaving aside more institutional concerns (e.g. about the appropriate division of power between judges and prosecutors), the argument I have just sketched seems to me the best defense that can be mounted for giving judges broad discretion at sentencing. That said, there are other features of proportionality that weigh in favor of a more rule-like approach to sentencing. I have suggested representing proportionality judgments as a function linking permutations of case-level factors to sentences. Suppose that we list the n factors, $f_1, f_2 \dots f_n$, that bear on a proportional sentence. A factor that is not included on that list is irrelevant and should not affect our judgment as to what

sentence is deserved (the defendant's hair color, for instance.) On the other hand, n must be finite—even defenders of discretionary sentencing must concede this point, since if a sentencing judge had to contemplate an infinite number of factors before rendering a sentence, she would never get around to sentencing anyone. However, being finite does not mean that the list is closed. We could always come to realize that we had left a relevant factor off the list, so the value of n could always turn out to be larger than we had previously thought.

A natural understanding of the proportionality function is:

Supervenience: proportionality judgments supervene on the set $[f_1, f_2 \dots f_n]$.

This means that if it is the case that a proportionate sentence for Jack is different from a proportionate sentence for Jill, there must be a difference between the set $[f_1 \text{ Jack}, f_2 \text{ Jack} \dots f_n \text{ Jack}]$ and the set $[f_1 \text{ Jill}, f_2 \text{ Jill} \dots f_n \text{ Jill}]$. By the same token, if Jack's situation is identical to Jill's along all n dimensions, then proportionality should recommend the same sentence for both. More colloquially, I am assuming that what makes a given punishment proportionate in one case can be explained by a combination of factors, such that, should the same combination of factors arise in a different case, they would explain why the same punishment would be proportionate in *that* case. By extension, if you think that what is proportional for Jack is different than what is proportional for Jill, then it must be for some articulable reason that serves to distinguish the two cases.

It is worth stressing that supervenience does *not* presuppose that the

proportional sentence can be determined by summing up the discrete contribution of each salient moral property. Supervenience operates across sets of properties, and hence is consistent with holism. I do not claim that *every* reasonable conception of proportionality necessarily incorporates supervenience. Perhaps there are other conceptions that are both plausible and inconsistent with it. However, I am inclined to think that supervenience is a bit hard to resist. To deny supervenience is to hold that two cases can be exactly alike in all relevant respects yet merit different punishments on grounds of proportionality. It is to claim that the constellation of reasons that make a given punishment proportionate in one case would not necessarily justify a similar punishment in another case exhibiting the same constellation of reasons. Perhaps some moral particularists would be prepared to defend these claims, although I won't explore the issue here.¹³

Supervenience seems especially hard to resist when what we are concerned with is not private judgments about what various people—friends, family members, co-workers, celebrities—deserve for their various transgressions, but rather with a system of public law governing how officials and institutions treat people. A system by which institutions could impose substantially different treatments on identically situated people would seem to open the door to intolerable arbitrariness. It would seem to foster social relations of subservience between citizens and officials, for citizens would be in no position to gainsay the judgment of officials. A citizen would have no basis for complaining that she was being treated

differently from how someone else had been treated. Consider the concerns about systematic bias in the criminal justice system I mentioned earlier. Rejecting supervenience would make it difficult to even understand those concerns, which are concerns that people alike in all relevant respects are being treated differently by public institutions. To deny supervenience is precisely to deny that people who are alike in all relevant respects are entitled to equal treatment.

As I understand it, proportionality is an individualized business, in that it is sensitive to the factors particular to each person's case. However, because it is sensitive to those factors in a uniform and predictable way, then so long as this standard of proportionality is uniformly applied across cases, the resulting set of punishments will be both fair and non-arbitrary. That the punishments will be non-arbitrary is easy to see: they will be non-arbitrary because each person's punishment will be a function of the factors that are constitutive of proportionality, and no others. What about fairness, in the sense of treating like cases alike? A system of individualized, proportional sentencing—at least if it satisfies supervenience—will generate a pattern of outcomes that satisfies a plausible construal of fairness. Because each sentence will be a function based on the same set of relevant factors, and no others, the sentences imposed will satisfy anonymity:

Anonymity: a proportionality judgment is anonymous if and only if it is insensitive to who commits the crime.

A system of punishment that satisfies supervenience will also satisfy anonymity, because if the values of $[f_1 \text{ Jack}, f_2 \text{ Jack} \dots f_n \text{ Jack}]$ do not differ from the values of $[f_1 \text{ Jill}, f_2 \text{ Jill} \dots f_n \text{ Jill}]$, then supervenience requires that they be treated equivalently. In other words, if Jack and Jill commit the exact same crime, under the exact same circumstances, then the punishment that is proportionate for one is proportionate for the other. That represents a recognizable and substantive account of fairness in punishment.¹⁴

This brief discussion suggests that, when it comes to proportional sentencing, non-arbitrariness and fairness are compossible virtues. Debates about structured versus discretionary sentencing often give the impression that we must choose between individualized sentencing and fair sentencing, but this is a misleading impression. In principle, we do not need to choose between sentencing that is tailored to the individual case and sentencing that is fair to everyone. We can endorse both values, and without preference for one over the other. The difficulty, of course, is in figuring out how to design a sentencing practice that—in light of the enormous complexity involved in holistic moral judgment—adequately achieves both non-arbitrariness in individual cases and systematic fairness.

2.

I have just suggested that acknowledging holism in proportionality judgments is not logically inconsistent

with a commitment to systemic fairness. The problem is a practical one: how can we operationalize

proportionality so as to adequately respect both of these values? Discretionary sentencing regimes focus on the individual case to the exclusion of fairness. Strict mandatory sentencing grids focus on treating like cases alike but give insufficient attention to moral nuance in the individual case.

It is here that emerging technologies hold some substantial promise. The idea, in a nutshell, is to develop a sentencing algorithm that draws correlations between input variables of various kinds—reasons judges give for the sentences they impose, for instance—and outcomes, that is, concrete sentences. This could be done by means of traditional regression techniques applied to a large enough, and rich enough, data set of cases and case outcomes.¹⁵ Alternatively, and perhaps more promisingly, a machine-learning algorithm might be deployed to discover correlations on its own. A significant advantage of a machine-learning algorithm is, assuming new judgments are fed in, that it is self-updating. Unlike more static efforts to systematize past practice in the form of a sentencing grid, or to otherwise provide comparative information to judges about the behavior of their peers, a machine-learning algorithm revises its predictions in light of new evidence, in this case proportionality judgments that vary significantly from the algorithm's initial prediction.¹⁶ This would allow the algorithm to track judicial opinion automatically, without requiring continual intervention on the part of legislatures or sentencing commissions.

Another advantage of a machine learning approach is that, provided the feature set in the input data is rich enough, the algorithm does not need to be encoded with a theory of

proportionality. Instead, it learns correlations of its own between input features and outcomes. Assume that we have a large number of cases in which sentencing judges have determined what range of punishment would be proportionate. From those cases, it might be possible to acquire the information required to map constellations of inputs (the set of factors, $[f_1, f_2 \dots f_n]$) onto outputs [proportional sentences]). Moreover, it might also be possible to determine which factors (including potentially unacknowledged ones) are correlated with changes in outcomes, and which are not.¹⁷

A requirement for such an algorithm to “learn” correlations is a sufficiently rich dataset about the cases coming before judges, as well as the sentences that they ultimately impose. Acquiring this kind of data is a non-trivial matter, in part because systematic and rich information of this kind is typically hard to come by in criminal justice, and in part because given what is at stake in the criminal process, the incentive for the parties to manipulate, explain away, exaggerate, or obscure relevant information is high. However, precisely because the stakes are high, the moral imperative to minimize arbitrariness in a sentencing system is correspondingly more urgent. That public institutions either have not collected information in the past, or have not done so systematically and transparently, should not be an excuse for insulating the status quo, particularly when it is evident that public institutions have for some time failed to treat people fairly.

Supposing we had access to the kind of data needed to train a proportionality algorithm, what the algorithm would essentially do is provide

sentencing judges with a particularized snapshot of the central tendency of how they and their colleagues have been treating similar cases. Judges who are provided with that information might be inclined to deviate less from the norm, absent truly idiosyncratic features. This would reduce variation among judges, but because it would do so on the basis of judges' *own* sentencing practice—what they and their colleagues have regarded as proportionate—it could not be said that this change in the distribution of sentences would be brought about by ignoring proportionality. After all, it would be based on their own collective wisdom, and drawn from the same sources that are used in rendering sentencing decisions as it stands. Insofar as the correlations are robust, the algorithm would be no worse at predicting proportionality than a discretionary system of sentencing, at least on average.

The predictions made by the algorithm should be treated as just that—predictions. Consequently, they should not be binding on sentencing judges, at least initially.¹⁸ The algorithm's predictions would serve merely to inform judges of what has been deemed proportionate in cases exhibiting a similar constellation of relevant factors; judges could set aside the algorithm's recommendation if they deem the case before them to be highly unusual. Indeed, judges should be encouraged to set aside the algorithm's prediction in unusual cases, with reasons explaining in what respect the case is unusual. Those judgments, in turn, could be used to further enrich the algorithm's predictive capacity going forward. Requiring reasons would help ensure that sentencing does not

become merely an exercise in rubber-stamping, as a judge would be required to consider whether the case before her is routine (in which case the algorithm's prediction is more likely to be accurate) or unusual in some significant respect (and hence requiring greater deliberation). That said, some scholars have doubted whether a purely advisory prediction will have much effect.¹⁹ The evidence is disputed, but if necessary, the algorithm's predictions could be treated as presumptively binding, with departures appealable by either party. This might further ensure adequate attention to case-specific features at sentencing.

Before proceeding, it is important to clear away an objection. Part of what makes sentencing look like a black art is that the inputs and outputs are so qualitatively different. On the input side, we are faced with a rich moral tapestry of actions, intentions, emotions, harms and relationships. These have to be translated, somehow, into a specific term of years—ultimately, a number.²⁰ One might think that because it is so unclear how this is done, it is even more unclear that a machine could replicate it in a convincing way. Hence, one might be inclined to be skeptical that a machine, however sophisticated, could really come to understand and apply the concept of proportionality to actual cases.

This objection rests on a fundamental misunderstanding. It is no part of the aim of a sentencing algorithm, at least of the kind I am envisioning, to replicate the kind of reasoning that human judges go through in pronouncing a sentence. What it is intended to do, rather, is to provide a reliable and accurate prediction of what a typical

sentencing judge in the relevant jurisdiction would regard as proportionate on a given set of facts. Consequently, the algorithm does not seek to model the moral content of proportionality as it figures in moral reasoning. Instead, it seeks to predict outcomes related to how human judges apply that concept in practice. Consider an analogy: algorithms have been devised to predict peoples' taste in music or films. This does not imply either that people develop a sense of taste in music or film by applying general rules of the kind observed by the algorithm, nor that the algorithm must model the way in which people do develop their tastes in order to successfully predict their taste in music or films. The sentencing algorithm stands to proportionality judgments in the way that a music or film-recommending algorithm stands to taste. Its criterion of success is how well it *predicts* the latter, not how well it *applies* it.²¹

None of this shows, of course, that it will be possible to construct an algorithm to connect the relevant features of a criminal case to a judgment about what kind of punishment is proportionate, but it does undermine the thought that a sentencing regime must rely on the unstructured exercise of human judgment in order to respect proportionality, either because the range of potentially relevant variables is indefinitely large, or because they can be combined in complex ways. What is required is some way of mapping distinct permutations of the factors that we have so far acknowledged to be morally relevant to punishments. While such a mapping would surely be very complex, in large part because of the number of potentially relevant factors, it is not clear that it would be beyond our abilities to represent that mapping in the form of a finite algorithm.

3.

The account of proportionality I have been sketching can be made intuitive by the following thought experiment.²² Suppose one hundred different judges are given the same set of facts and are asked to pronounce a proportional sentence. We might expect to get back a range of answers clustered around a mean. Perhaps the answers would be distributed normally, as in Figures 1 and 2.

Note that these figures represent a distribution of judicial opinions as to the proportional sentence for a *specific* case, rather than a *class* of cases (e.g. armed robberies or break and enters.) A graph of sentences over a class of cases may have a similar

distribution, but variation in that graph will, among other things, be responsive to variation among the instances that make up the class of cases, for instance, the proportion of cases that are more rather than less aggravated in nature. In contrast, when we are looking at a distribution of opinions about the proportional

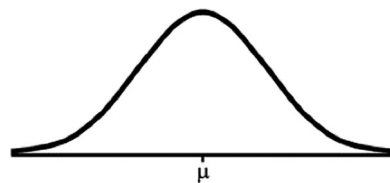


Figure 1. Sentences widely distributed around a mean.

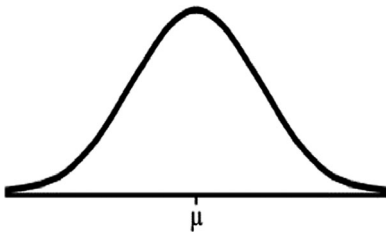


Figure 2. Sentences more narrowly distributed around a mean.

sentence on a single case, then the variance, σ^2 , represents our commitment to the proposition that like cases should be treated alike.

Arbitrariness in sentencing can be described in terms of the variance in the distribution. By the same token, variance in the distribution is a measure of the degree to which a sentencing system (in this case, discretionary sentencing) realizes supervenience. After all, the proportionality judgments lying along the curve are all outputs of the same set of factors, $[f_1, f_2 \dots f_n]$. If proportionality judgments supervened perfectly upon those factors, then one would expect the judgments to converge, e.g. at μ . However, some arbitrariness is likely to be inevitable, as our judgments about proportionality, culpability, etc. are far from precise, and it would be unrealistic to expect consensus as to what specific sentence is proportional for a given case, although one might expect consensus on a broader range within which the sentence should fall. Supervenience requires, therefore, that like cases be treated alike in the sense that they fall within some acceptable range—the variance of the distribution should be finite and reasonably tailored—but it does not require convergence upon some specific sentence.²³

Of course, it would be difficult to specify the appropriate level of

variance in numerical terms. More plausibly, however, one can proceed qualitatively. Some sources of variance are impermissible, whether they pertain to the accused (attractiveness) or to the judge (mood, or proximity to an election). Sometimes these can be a matter essentially of chance (whether your case is heard before or after the lunch break), but sometimes it can reflect more systematic biases. For instance, some judges may have (conscious or implicit) biases for or against certain demographic groups. Bias of this latter kind violates anonymity. It suggests that your chance of ending up on the left or right side of the distribution is affected by who you are, not just by what you did.

The total variance in a distribution of outcomes is likely to be combined from both (a) justified variance due to legitimate differences of opinion or vagueness in our moral concepts and (b) unjustified variance that reflects either impermissible bias or random noise. The debate about discretionary and structured sentencing can be understood as centered on the question of which sentencing regime accommodates (a) while minimizing (b). Defenders of discretionary sentencing might be interpreted as arguing, in effect, that no other regime would allow judges to react appropriately to all the morally salient factors as they pertain to a single case (accommodating (a)). Critics, in contrast, emphasize the failures of discretionary sentencing in controlling (b).

In light of this discussion, how would a sentencing algorithm of the kind I have envisioned change a system of discretionary sentencing? It should have the effect of making the distribution somewhat narrower (reducing variance); it should, in other words, encourage a distribution

more like Figure 2 than Figure 1. If the algorithm accurately reads past cases, it would predict μ on the given facts. However, it would not seek, nor result in, convergence. A judge may depart from μ if she deems it morally appropriate under the circumstances. This departure from μ would be taken into account in updating the algorithm, improving the predictive accuracy in future iterations.

The moral value of substituting a narrower distribution for a broader one turns on the plausibility of two propositions. First, the variance in the distribution of sentences that would be produced by existing discretionary sentencing regimes flows from both justified vagueness/differences of opinion as to proportionality as well as unjustified (perhaps implicit) bias and random noise. Second, a sentencing algorithm can help distinguish between the two. For instance, suppose that hearing a case earlier in the day tends to push a judge toward the left end of the distribution, whereas hearing a case later in the day tends to push a judge toward the right end. If cases are equally likely to be heard in the morning as in the afternoon, then the algorithm's prediction of μ would in effect average out the impact of time of day on the sentence. Consequently, a judge relying on the algorithm would see the same μ whether she was sitting in the morning or in the afternoon. Insofar as the judge's inclination to depart from the algorithm is not itself determined by time of day, we would expect to see

the distribution become narrower, not because the judge is prevented from considering morally relevant considerations, but because of steps to constrain the influence of morally irrelevant ones.

This line of argument suggests that a sentencing algorithm should satisfy both defenders and critics of discretionary sentencing. It should satisfy defenders because it does not involve artificially restricting a judge's considerations of the morally relevant factors in a given case; indeed, the algorithm's predictions are built on the prior practice of judges, thereby reflecting the particularized judgment that defenders of discretionary sentencing are keen to stress. On the other hand, by systematizing that past practice, the algorithm would help minimize unjustified arbitrariness. Outliers would be pulled in closer to the mean by revealing to judges when they are far from the mean, and would thereby help limit the impact of random noise in sentencing. Consequently, insofar as the algorithm exerts influence on sentencers, it would have the effect of pulling in the tails of the distribution. This would make sentencing more predictable and less arbitrary. It would do so by making the sentence a person receives more fully a function of the morally relevant factors that apply to his case (as demanded by supervenience.) I suggested earlier that non-arbitrariness and fairness are compossible virtues. A sentencing algorithm illustrates how that can be so.

4.

Although using an algorithm to predict proportionality may seem novel, many jurisdictions have been

trying, for quite some time, and with quite a number of variations, to use other means to achieve the same

end. Most notably, this proposal updates earlier efforts to reduce arbitrariness in sentencing by encouraging judges to use systematically collected information about sentencing practice among their colleagues, information that could be accessed and represented in a computerized form. Those efforts, notably in Canada in the 1980s and Scotland in the 1990s, operated on the basis of a database of cases that could be searched for “similar” cases along a range of dimensions (criminal history, crime of conviction, etc.), and which would then provide information regarding the distribution of sentences for cases in that class. Those efforts failed, largely because of faltering political and institutional support, as well as a lack of “buy in” by judges, who were often unwilling to depart from the subjective, case-by-case, legalistic modes of sentencing to which they were accustomed.²⁴

My concern in this paper is whether such an algorithm is desirable in the first place, rather than the hurdles facing implementation, although those hurdles are admittedly substantial. (The experience in Canada and Scotland suggests that the most significant barriers are likely to be political and social, rather than technical.) There are two reasons for revisiting this issue. First, the impressive, and rapidly accelerating, improvements in predictive analytics over the last two decades—from self-driving cars to predicting taste in movies and music to outperforming human champions at chess—suggests that a sentencing algorithm may be able to provide much more accurate and granular predictions than in previous generations. Second, concerns about racial

inequity in criminal justice have become increasingly prominent in that same time period. Traditional views of sentencing as inherently subjective and unquantifiable are in some substantial tension with these concerns. As with the broad discretion of police officers to arrest or detain individuals, broad discretion at sentencing can serve to mask unconscious, but nevertheless inequitable, practices. While a sentencing algorithm cannot be expected to be a complete remedy for discriminatory sentencing, minimizing the unjustifiable variance in imposed sentences is, for reasons discussed below, an important partial step toward that end.

Looking beyond computerized sentencing aids, one might think of a proportionality prediction as a case-specific application of the first two steps of sentencing as established by the definitive guidelines set out by the Sentencing Council for England and Wales. Under that framework, a sentencing court is asked to first establish the category in which a given crime falls by ranking the crime according to the degree of harm inflicted as well as the offender’s culpability. In the second step, the sentencing court is referred to a range of sentences for that category of offence as well as a “starting point” within that range. The court is then required to take into account a non-exhaustive list of case-specific aggravating and mitigating factors, adjusting the starting point accordingly. The remaining steps modify the sentence for systemic factors, such as early guilty pleas, time spent in pre-trial detention, and so forth.²⁵

Although it does not involve guidelines or starting points, one

might also think of recent amendments to the law of sentencing in Israel as falling along these lines. The Israeli Penal Law requires a two-step sentencing process to give effect to proportionality. First, a sentencing judge must establish a “Proportionate Sentence Range” for a given case, taking into account the seriousness of the crime and the culpability of the accused. Second, the judge must settle upon a specific sentence within that range on the basis of an enumerated list of factors, the accused’s efforts to make amends, life circumstances and so forth.²⁶ By requiring a sentencing judge to state a proportional range of sentences, the law seeks to ensure some level of systematic fairness even as, in the second step, the judges are able to tailor a sentence to the individualized circumstances of the accused person before them.

These are just a few examples of efforts that different countries have developed for trying to adequately respect both individualized sentencing and systematic fairness. A predictive algorithm of the kind I am proposing is thus not *sui generis*. It

represents an incremental development in well-established trends in sentencing in a number of jurisdictions. There are two main differences between a predictive algorithm and more traditional approaches. The first is that the algorithm bypasses the categorical first step (establishing the class of offenses for which a range of proportional sentences is defined) and goes straight to predicting a proportionate sentence for the individual case. The kind of uniformity that is sought by requiring judges to apply grids, starting points, or proportionality ranges is instead built into the correlations that the algorithm learns from the data fed into it. The second is that a machine-learning algorithm learns from experience, in that it updates its predictions in light of the success or failure of its prior predictions. Cases in which judges depart significantly from its predictions would provide a basis for updating its predictions going forward, ideally rendering the predictions both more accurate and up-to-date with changes in judicial opinions about proportionality.

5.

Unlike in the context of pre-trial detention, where a variety of risk-assessment algorithms have been devised and implemented, there is much less consensus about the purposes of sentencing. I have been focusing on proportionality, but of course ensuring that punishments are proportional is not the only thing we care about in sentencing.

In fact, it is rather hard to say what the point of sentencing is meant to be. In Canada, for instance, §718 of the

Criminal Code states that the “fundamental purpose of sentencing” is

To contribute, along with crime prevention initiatives, to respect for the law and the maintenance of a just, peaceful and safe society by imposing just sanctions that have one or more of the following objectives: (a) to denounce unlawful conduct; (b) to deter the offender and other person from committing offences; (c) to separate offenders from society, where necessary; (d) to assist in rehabilitating offenders; (e) to provide reparations for harm done to victims or to the community; and (f) to provide a sense of responsibility in offenders,

*and acknowledgment of the harm done to victims and to the community.*²⁷

The multiplicity of purposes described in §718 would be less of an issue if they all tended to move in the same direction. Unfortunately, however, §718 is a mere laundry list: the items on the list bear no obvious relation to each other and can often be expected to pull in different directions. It is hard to see, for instance, how providing reparations to the victim is advanced by incarcerating the accused. Some serious crimes might be effectively undeterrable, but there might still be substantial reason to condemn them. Incapacitating someone who we believe is likely to commit more crimes in the future may be unnecessary to acknowledge the harm he caused to his victim, and so forth.

What this means is that it can be very difficult to say what exactly it is that we, as a society, are trying to achieve by punishing criminals. Perhaps this is unsurprising. Sentencing reform has been, and continues to be, a hot-button political issue in Canada, particularly when it comes to making difficult choices in how to prioritize among the different possible objectives we might want sentencing to achieve. As a result, as Julian Roberts has noted, even setting up a permanent sentencing body that could approach sentencing reform in a principled manner has proven to be a bridge too far.²⁸ Despite significant efforts to rationalize a highly discretionary sentencing model in the 1980s, sentencing “reform” in Canada has amounted to little more than a recapitulation of the status quo of highly discretionary sentencing. Because §718 merely requires a sentence to advance “one

or more” of the subtended objectives, and because the objectives are often at odds with each other, the statute cannot be said to even aim at ensuring that like cases are treated alike.

That said, proportional sentencing nevertheless remains a popular idea.²⁹ Indeed, §718.1 of the Criminal Code gives proportionality a special status by requiring, without qualification, a sentence to be “proportionate to the gravity of the offence and the degree of responsibility of the offender.”³⁰ Taken together, the message of §718 and §718.1 is: we are not sure what we are trying to do by punishing criminals, but whatever it is, we should only do it by means that are proportionate.

This suggests that making proportionality judgments less arbitrary and more uniform will be valuable even when there is significant disagreement as to what we are trying to achieve in sentencing, and even if a proportionality judgment is merely the starting point for a sentence rather than its ultimate disposition. True, there is far less consensus about the purposes of punishment than there is about the purposes of pre-trial detention, and, in light of that incoherence, it would be challenging to specify some more general function mapping crimes onto sentences. However, there is a high degree of consensus, as well as (in jurisdictions like Canada) a legal mandate to prioritize proportionality. Hence, a more restricted algorithm that provides guidance on what proportionality requires in each case would be of substantial value. For instance, a sentencing judge could be asked to take a predicted sentence as a starting point, and to explain departures

from it on the basis of one or more of the purposes of punishment noted in §718, so long as those

departures are not so substantial as to render the ultimate sentence disproportionate.

6.

What could possibly go wrong? Here are some thoughts.

First, one might worry that an algorithm would reproduce discriminatory outcomes created elsewhere in the system. Black kids typically have more criminal justice contacts than white kids. Insofar as criminal history bears upon proportionality, it might seem that an algorithm would simply reproduce discriminatory treatment from elsewhere in the system. Many have observed that if you go looking for crime, sooner or later you will find it, meaning that contact with the criminal justice system is not simply a function of participation in crime, but is also a function of policing and enforcement priorities.³¹

This type of concern is familiar from many contexts, including contexts that have nothing in particular to do with algorithms or machine learning. Is there particular reason to worry about this effect in the context of algorithmic decision-making?³² One reason that might be specific to algorithmic sentencing is that, by taking discretion out of the hands of sentencing judges, it would in effect “bake in” discriminatory decisions made elsewhere in the system, making them harder to change.

Here, it is important to consider: compared to what? As I have suggested, algorithmic sentencing need not be perfect. It just needs to be better than the status quo; that, unfortunately, is a rather low bar. Our existing systems of criminal

justice are marked by extreme and persistent racial disparities. This suggests that discrimination is *already* baked into our existing systems of criminal justice. These are systems built on people—police, lawyers and judges—exercising their clinical, case-level discretion, often with little by way of systematic oversight. Judges exercising broad sentencing discretion do not appear to be immune to the same biases that afflict human decision-making generally, including along the dimension of racial bias.³³ While conscientious judges will disregard factors they regard as morally irrelevant, this does not address unconsciously held biases. Human judges, like most humans, are subject to a range of biases, implicit or otherwise.³⁴ Such biases, moreover, are not necessarily remedied simply by bringing them to attention of those who have them. More generally, even the most well-intentioned discretionary sentencing system can sustain racially disparate patterns of punishment. Consider that the Supreme Court of Canada has long encouraged lower courts to give “particular attention” to the circumstances of indigenous offenders, in light of the overrepresentation of indigenous persons in Canadian jails and prisons.³⁵ Yet despite that well-known exhortation by Canada’s highest court, the overrepresentation of indigenous people in the criminal justice system has not only failed to improve, but by some measures has actually worsened.³⁶

Of course it is true that perfectly even-handed sentencing will not eliminate racial disparity in our criminal justice systems, but that is not the right standard here. The sources of that disparity are multifarious and complex, and often involve decisions made well in advance of sentencing. It seems unreasonable to expect that any change to the status quo in sentencing must be tasked with fixing problems that have nothing to do with sentencing. In any case, there is reason to think that algorithmic sentencing can actually make sentencing fairer than the status quo. Consider the case of risk assessments in the pre-trial context. Kleinberg et al. have recently noted, on the basis of a large study of release decisions in New York, that a machine learning algorithm could simultaneously reduce both overrepresentation of racial minorities and failures to appear in subsequent court proceedings.³⁷ Under the status quo, judicial decisions about whether to detain someone pending trial are a source of further racial skew: although 82% of arrestees are black or Hispanic, nearly 90% of detained people are. Kleinberg et al. show that the proportion of minority people detained could be made to mirror the proportion of arrestees—thereby taking pre-trial detention out of the loop as an independent source of racial bias—while still decreasing future crime by nearly 23%.³⁸

Could a sentencing algorithm of the kind I am envisioning lead to similar results in the sentencing context? It would, for instance, if it contributed to reducing the impact of implicit bias by sentencing judges, either premised upon racial classification itself, or upon features that are closely correlated with race. Of

course, a sentencing algorithm would have to be designed in a way that is sensitive to the likelihood that existing judicial sentencing practice already reflects racial bias. How best to do this is a matter of ongoing discussion among computer scientists and lawyers.³⁹ The algorithm might, for instance, be made to predict similar outcomes while controlling for race and features strongly correlated with race. More aggressively, the algorithm might be constrained to ensure roughly equal rates of both false positives and false negatives; or, more abstractly, to minimize the net burden upon racial minorities.⁴⁰

Consider, again, the narrower distribution of sentences (Figure 2) that I suggested would be the likely upshot of a sentencing algorithm. I suggested that the narrowing could be defended on grounds of separating out justified from unjustified sources of arbitrariness in sentencing. In principle, however, that could happen even while leaving the relative position of black and white defendants unchanged, that is, with black defendants disproportionately falling on the right tail of the distribution. However, if a sentencing algorithm reduced the impact of unjustified arbitrariness correlated with race, then it would simultaneously make the distribution narrower (reducing variance) and even out the disparity in the likelihood that a black rather than a white defendant will end up on the right tail.

That said, suppose that a sentencing algorithm does not correct racial disparities produced earlier in the criminal justice process. There would nevertheless be a significant, if incomplete, advantage to racial fairness simply by virtue of its effect in narrowing the variance in the

distribution of outcomes. Decreasing variance means that sentencing disparities become smaller in absolute terms, even if their relative position remains unchanged. That would be a significant achievement in its own right, even though it is by no means a solution to racial inequity in sentencing. Suppose a biased judge sentences a black defendant to eighteen months, whereas she sentences an identically situated white defendant to six months. It would be an improvement if she instead sentenced the black defendant to fifteen months and the white defendant to nine months, even if it were better still if there were no bias in the first place. Incarceration is a liberal society's most serious intentionally inflicted deprivation. Reducing the degree to which its use is infected by unjustified random arbitrariness improves the situation of racial minorities, who are disproportionately represented in custodial populations, even if it is not a complete answer to unjustified racial arbitrariness.

A second source of worry has to do with accountability. At least if you are unhappy with how you have been treated by a sentencing judge, you have someone to blame, but if you are unhappy with how you have been treated by an algorithm, there is no one to blame. The other actors in the system—police, prosecutors, judges—can evade responsibility by saying, in effect, “the algorithm made me do it.”

This worry strikes me as easier to answer than the first. Algorithmic sentencing recommendations do not remove accountability, and may actually enhance it. Firstly, on the model I am envisioning, the algorithm is non-binding, and a sentencing judge would have to give due consideration

to whether its estimate of proportionality is appropriate in a given case.⁴¹ Hence, judges could not evade responsibility by saying that the algorithm made them do it. The algorithm would not “make” anyone do anything.

Algorithmic sentencing may enhance sentencing accountability by ensuring that the basis for sentencing judgments is easier to understand and evaluate. Although sentencing judges obviously provide reasons for their sentences, how those reasons translate into a specific sentencing recommendation is mysterious, and often hard for outsiders and non-specialists to evaluate. It is hard to know what to make of a specific sentence, unless you know a lot more about what other people received in more or less analogous situations, as well as what other people, facing different charges and in different contexts, received. It is hard to evaluate the proportionality of a proposed sentence without knowing a lot about the prevailing sentencing practice in a given court.

With an algorithm, by contrast, in principle anyone can see how varying this or that factor affects proportional sentences. This would make it much easier for the public to have an informed discussion about proportionality in punishment than the current status quo of discretionary sentencing. It is also easier to modify an algorithm than it is to change the minds of sentencing judges. The algorithm represents our collective wisdom about what kind of sentence is proportionate for the kinds of cases we typically have to deal with. If we change our mind about that, it is easier to implement that change by tweaking an algorithm than by persuading thousands of sentencing judges to change what they

may regard as settled practice. Errors are less sticky in algorithmic decision-making processes.

To be clear, I am operating on the assumption that the algorithm would be publicly reviewable and contestable, and that it would not be a proprietary product.⁴² While contestation could take the form of litigation brought by, or on behalf of an individual sentenced under such a system, litigation is not the only, or even necessarily the most meaningful, form of contestation. The competence to assess how a machine-learning algorithm is designed, trained, and validated falls well outside the expertise of most courts, and indeed, most laypersons generally. Consequently, a more meaningful form of contestation is promoted by ensuring that the algorithm remains accessible to independent researchers and experts. One could, for instance, envision that a sentencing algorithm be required to undergo regular validation by a panel of such experts, with a mandate to review the technical parameters of the algorithm's design, as well as to ensure that its predictions fall within acceptable tolerances. Beyond expert-driven, institutional oversight, individual cases that raise truly unusual circumstances (as in the case of the father who ends the life of his severely disabled daughter, discussed above) would be grounds for setting aside the algorithm's predictions. Sentences would remain appealable, and modifications on appeal should be used to train the algorithm to provide increasingly accurate predictions. Admittedly, this would require the support of a jurisdiction's appellate courts, which cannot be taken for granted.

I see no reason why access to the algorithm's predictions should be

restricted to sentencing judges. The algorithm, perhaps in the form of a website or mobile app, would allow all the parties—defense counsel, prosecutors, victims and, crucially, the accused themselves—to estimate the probable sentence if convicted after trial. By varying the inputs, the parties could also derive a good sense of how material winning or losing on a particular issue might be, whether at trial or at sentencing. For instance, the parties might realize that instead of focusing on motive they would do better to focus on the harm suffered by the victim, or vice versa. Thus, just as the reliance on a predictive algorithm may affect the variability of sentences, it may similarly affect the nature of sentencing proceedings.

Moreover, as it stands, most criminal accused depend upon their lawyer—assuming they have one—to give them an accurate estimate of the punishment they can expect to receive should they be convicted after trial. But since the factors that predict proportionality are unlikely to be particularly obscure, legalistic, or otherwise require professional advice to interpret, a proportionality algorithm may tend to level the playing field between well-resourced accused who are able to retain experienced defense counsel and indigent accused who have far less access to accurate and informed legal advice. This suggests that a predictive algorithm would empower criminal accused, with the bulk of that benefit accruing to those with the least access to competent counsel.⁴³

One might be tempted here to give in to an anti-populist sentiment, namely the belief that once the public can see what influences sentencing, they will be tempted to change

it, and not for the better. Criminal justice is highly susceptible—particularly in the United States, but to some degree in Canada, too—to demagogic politicking. Once the public knows what is going on, they might be upset and agitate for change. One might therefore be inclined to resist doing anything that would render the sentencing process more publicly observable.

For my part, I think we should resist this sentiment. It speaks poorly of a government with democratic pretensions if it only feels comfortable punishing its citizens so long as those citizens can be relied upon to remain

ignorant of the reasons why they are being punished to this or that degree. This is not to say that sentencing decisions should be a matter of popular referenda. Making sentencing more publicly observable does not imply making it subject to unfiltered popular control. Precisely because sentencing is so momentous, there is probably good reason to insulate it even further from popular sentiment than is currently the case, for instance through appointed, non-partisan and expertise-driven sentencing commissions. That is consistent, however, with making the grounds of punishment more publicly observable.

7.

There is evidence that people generally overestimate their ability to outsmart well-designed algorithms, and there is also evidence that discretionary judgment is correlated with bias of various kinds.⁴⁴ Criminal sentencing is without a doubt difficult. It is hard to believe, however, that it is so difficult that it could not possibly be improved upon by an algorithm, no matter how advanced. That claim has been made before, in a wide variety of contexts. It has repeatedly turned out to be false. Maybe it will turn out to be false here, too.

One hundred and fifty years ago, Darwin unsettled many people with

his suggestion that humans might be descended from primates. While most people today have come to terms with that idea, many are still confident that there are activities that are uniquely and irreducibly the domain of humans. The moral judgment involved in sentencing a person for a crime sometimes seems to be one such activity. Hence, the thought that an algorithm—manifested in a website or a mobile app—might one day be able to reproduce that kind of judgment can be deeply unsettling. But it also, or so it seems to me, flows from Darwin's radical idea.

Notes

[**Disclosure Statement:** No potential conflict of interest was reported by the author.]

1 Unwarranted sentencing disparity was at the heart of a split judgment in a recent Alberta Court of Appeal judgment. See *R v Ryan*, 2015 A.B.C.A. 286 (Can. Alta. C.A.).

2 See American Law Institute, Model Penal Code: Sentencing (Tentative Draft), § 6B.09 (“Evidence-Based Sentencing; Offender Treatment Needs And Risk of Reoffending”), §§ (2) and (3).

3 There is a large body of literature on risk-assessment devices, both supporting and critical; see, for example, Harcourt, *Against Prediction*; Starr, "Evidence-Based Sentencing"; Hannah-Moffat, "Actuarial Sentencing"; Chanenson and Hyatt, *The Use of Risk*.

4 Others have made similar proposals. See, for instance, Adi Leibovitch's argument in favor of "curving" sentencing discretion by providing judges with statistical information about how similar cases are treated in different courts in order to ensure that harmony between sentences is imposed in specialized and generalist courts ("Punishing on a Curve"). In a similar spirit, Laqueur and Copus propose a "synthetic crowdsourcing" approach to resolving inconsistency in parole hearings (see "Synthetic Crowdsourcing"). More generally, Bagaric and Wolf raise many of the points considered here in their discussion of computerized sentencing (see "Sentencing By Computer").

5 Similar tools are already in existence. For instance, a Canadian firm has developed a searchable database to predict sentencing ranges based on decided cases that share similar features to a given case. See <http://www.rangefindr.ca/>

6 As a result, my proposal is not an instance of what Harcourt would regard as an "actuarial" method, as he defines the term. For Harcourt, "actuarial methods" are aimed at predicting "past, present or future criminal behavior of a particular person" (*Against Prediction*, 16). My proposal is focused on predicting judicial opinion, not criminal behavior.

7 See e.g. Morris, *The Future of Imprisonment*.

8 See Frase, "Limiting Retributivism," 93.

9 More broadly in the philosophy of criminal law, the distinction reflects a divide between more Kantian approaches to punishment, which insist on punishment to vindicate purely abstract rights; and more liberal approaches (associated most prominently with H.L.A. Hart), which insist that punishment is legitimate only insofar as it furthers valuable social objectives.

10 See *R v Latimer*, [2001] 1 S.C.R. 3 (Can.).

11 *Ethics Without Principles*, 7. Moral particularism is typically, although not inevitably, associated with holism.

12 For a helpful discussion of holism in the context of moral particularism, see Little, "Moral Generalities Revisited." The significance of holism in this context has not gone wholly unnoticed; see Tata, "The Application of Judicial Intelligence."

13 It may be worth noting here that supervenience does not imply that there are patterns or general principles linking the facts of a case to proportionality judgments. It only entails that, as Margaret Olivia Little puts it, "[t]wo situations... cannot differ in some moral respect without differing in some nonmoral respect" ("Moral Generalities Revisited," 280–81); see also Lance and Little, "From Particularism to Defeasibility"; Dancy, *Ethics Without Principles*, 85–88.

14 Of course, these features might themselves make an issue of the criminal's identity—for instance, that the accused was a family member. But that does not violate anonymity, for the same punishment will be given to anyone who commits a similar crime under similar circumstances, where those circumstances include standing in a familial relation to the victim.

15 Some of the risk assessment devices in the bail context operate in this manner. For instance, the Arnold Foundation's risk assessment tool for bail is based on a simple regression.

16 My thanks to Ryan Liss for emphasizing this point to me.

17 I am, obviously, not proposing any particular algorithm here. That said, it would be quite surprising—and rather disturbing—if it were to turn out that there are no meaningful predictive correlations between fact patterns and sentences imposed.

18 Making the algorithm's predictions merely advisory would help offset some of the widely noted problems with prosecutorial charging discretion in structured sentencing regimes, as judges would not become boxed in by how a prosecutor has chosen to structure an indictment.

19 See Tonry, *Sentencing Matters*. However, others have argued that the

evidence in favor of presumptive over advisory regimes is less clear cut, as some advisory regimes have compliance rates similar to presumptive regimes. See Hunt and Connelly, "Advisory Guidelines" and Reitz, "Comparing Sentencing Guidelines." Reitz observes that "[s]ome advisory guidelines have proven to be irrelevancies," although in other jurisdictions, "advisory guidelines have won a stature not so different from their presumptive counterparts" (196).

20 For simplicity's sake, I shall be focusing on custodial sanctions. Most of the attention on proportionality and parity in sentencing has focused on custodial sanctions as well, although they comprise the minority of criminal sentences.

21 This distinguishes my approach from others, such as that of Bagaric and Wolf, who would prefer to encode substantive sentencing principles into the algorithm directly:

We recommend that a constant, unvarying suite of factors that inform penalty, including aggravating and mitigating considerations that increase or decrease penalty respectively, and specifications of the weight that attach to each of those factors in certain circumstances, should be built into the computer algorithm. ("Sentencing By Computer," 33)

Because it presumes that concrete sentencing outcomes flow out of summing up aggravating and mitigating features, Bagaric and Wolf's strategy raises concerns in meta-ethics about the nature of moral judgment, concerns that the type of sentencing algorithm I envision avoids.

22 My thanks to Aziz Huq, who has helped me sort my thoughts in this section, and to whom I owe the representations in figures 1 and 2.

23 One could render this requirement as convergence upon the same *ex ante* chance of falling somewhere in the given distribution, with the understanding that the endpoints of the distribution are cabined within some fixed range.

24 See Tata, "The Application of Judicial Intelligence" and "The Struggle for Sentencing Reform," 247–48 (Scotland). Doob and Park, writing in 1987, presciently

observed that the success of sentencing reform depended upon judicial attitudes: "[i]f ... judges feel comfortable sentencing in the absence of systematic information about current practice, then no information system of any kind which provides this information is likely to be of use to them" ("Computerized Sentencing Information," 72). As Tata notes, judicial indifference ultimately doomed efforts to rationalize sentencing in Canada: see "The Application of Judicial Intelligence," 208–212.

25 See Coroners and Justice Act 2009, §§118–132 (Eng.), where the powers and responsibilities of the Sentencing Council are set forth. Under § 125(1), courts are required to adhere to the sentencing guidelines "unless the court is satisfied that it would be contrary to the interest of justice to do so." The existing definitive guidelines are available on the Sentencing Council's website: <https://www.sentencingcouncil.org.uk/>

26 See Penal Law (Amendment No 113) 2012, 2337 LSI 170 (Israel), § 40C; the enumerated factors are listed in §§ 40(d-e) and 40 (k). I am indebted to Julian Roberts for bringing this point to my attention; for discussion, see Roberts and Gazal-Ayal, "Statutory Sentencing Reform."

27 Criminal Code, § 718. The Supreme Court of Canada has added retribution to this list of purposes. See *R v CAM*, [1996] 1 S.C.R. 500 (Can.).

28 See Roberts, "Structuring Sentencing in Canada," 330.

29 In addition to the Canadian and Israeli sentencing provisions discussed above, the Model Penal Code describes the "general purposes" of its sentencing provisions as ensuring that sentences are "in all cases within a range of severity proportionate to the gravity of offenses, the harms done to crime victims, and the blameworthiness of offenders." See Model Penal Code, § 1.02 (2)(a)(i).

30 Criminal Code, §718.1. Similarly, the Israeli Penal Law describes proportionality as "the guiding principle in sentencing." Israel Penal Law, 5737–1977, § 40(b).

31 This is a theme in Elizabeth Hinton's recent book, *From the War on Poverty*.

32 Sonja B. Starr, for instance, has argued this point in the context of risk-assessment devices for pre-trial detention; see “Evidence-Based Sentencing.”

33 See Yang, “Free at Last?,” which finds evidence of greater racial disparities in the sentencing patterns of federal judges appointed after the Supreme Court made the Federal Sentencing Guidelines advisory, in *United States v. Booker*, 543 U.S. 220 (2005), than in the sentencing patterns of judges appointed earlier. Researchers have found that judges suffer the same kinds of implicit biases, including racial biases, as lay people. See Rachlinski et al., “Does Unconscious Racial Bias.”

34 There is a significant body of psychological literature on this subject; for an overview, see Englich, “Heuristic strategies”; Goodman-Delahunty and Sporer, “Unconscious influences in sentencing.”

35 See *R v Gladue*, [1999] 1 S.C.R. 688 (Can.).

36 See *R v Ipeelee*, [2012] 1 S.C.R. 433 62 (Can.).

37 See Kleinberg et al., “Human Decisions,” 29–32.

38 See Kleinberg et al., “Human Decisions,” Table 7.

39 See Huq, “Racial Equity”; Berk et al. “Fairness in Criminal Justice.”

40 See Huq, “Racial Equity.”

41 They would have to be alert to “broken leg” cases, that is, cases of improbable occurrences that, when they do occur, undermine even otherwise robust actuarial predictions. See Grove and Meehl, “Comparative Efficiency,” 307–8.

42 See *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016).

43 For discussion of some of the novel due process concerns raised by technological advances of this kind, see Citron, “Technological Due Process.”

44 See Grove and Meehl, “Comparative Efficiency”; Dana, Dawes and Peterson, “Belief in the unstructured interview,” which reports that unstructured interviews actually impair judgment; Bertrand and Mullainathan, “Are Emily and Greg,” which finds decisions about which job candidates to call back for an interview were biased against applicants with stereotypically African-American-sounding names relative to applicants with stereotypically white-sounding names, such that an applicant with an African-American name needed approximately eight additional years of experience to receive as many callbacks as an applicant with a white-sounding name).

Bibliography

- Bagaric, Mirko, and Gabrielle Wolf. “Sentencing by Computer: Enhancing Sentencing Transparency and Predictability, and (Possibly) Bridging the Gap between Sentencing Knowledge and Practice.” *George Mason Law Review* 25, no. 4 (Forthcoming). <https://ssrn.com/abstract=3011166>.
- Berk, Richard, Hoda Heidari, Shahin Jabbari, Michael Kearns, and Aaron Roth. “Fairness in Criminal Justice Risk Assessments: The State of the Art.” *Sociological Methods and Research*. July 2, 2018. <https://doi.org/10.1177%2F0049124118782533>.
- Bertrand, Marianne, and Sendhil Mullainathan. “Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination.” *American Economic Review* 94, no. 4 (2004): 991–1013.
- Chanenson, Steven L., and Jordan M. Hyatt. “The Use of Risk Assessment at Sentencing: Implications for Research and Policy.” *Working Paper Series* 193. Villanova University Charles Widger School of Law, Villanova, PA. December 2016. <http://digitalcommons.law.villanova.edu/cgi/viewcontent.cgi?article=1201&context=wps>.
- Citron, Danielle Keats. “Technological Due Process.” *Washington University Law Review* 85, no. 6 (2008): 1249–1313.
- Dana, Jason, Robyn Dawes, and Nathaniel Peterson. “Belief in the Unstructured Interview: The Persistence of an Illusion.” *Judgment and Decision Making* 8, no. 5 (2013): 512–20.
- Dancy, Jonathan. *Ethics Without Principles*. Oxford: Oxford University Press, 2004.
- Doob, Anthony N., and Norman W. Park. “Computerized Sentencing Information for

- Judges: An Aid to the Sentencing Process." *Criminal Law Quarterly* 30 (1987): 54–72.
- Englich, Birte. "Heuristic Strategies and Persistent Biases in Sentencing Decisions." In *Social Psychology of Punishment of Crime*, edited by Margit E. Oswald, Steffen Bieneck, and Jorg Hupfeld-Heinemann, 295–314. Hoboken: Wiley-Blackwell, 2009.
- Frase, Richard. "Limiting Retributivism." In *The Future of Imprisonment in the 21st Century*, edited by Michael Tonry, 83–119. Oxford: Oxford University Press, 2004.
- Goodman-Delahunty, Jane, and Siegfried Ludwig Sporer. "Unconscious Influences in Sentencing Decisions: A Research Review of Psychological Sources of Disparity." *Australian Journal of Forensic Sciences* 42, no. 1 (March 2010): 19–36.
- Grove, William M., and Paul E. Meehl. "Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical-Statistical Controversy." *Psychology, Public Policy, and Law* 2, no. 2 (1996): 293–323.
- Hannah-Moffat, Kelly. "Actuarial Sentencing: An 'Unsettled' Proposition." *Justice Quarterly* 30, no. 2 (2013): 270–96.
- Harcourt, Bernard. *Against Prediction: Profiling, Policing, and Punishing in An Actuarial Age*. Chicago: University of Chicago Press, 2006.
- Hinton, Elizabeth. *From the War on Poverty to the War on Crime*. Cambridge, MA: Harvard University Press, 2016.
- Hunt, Kim, and Michael Connelly. "Advisory Guidelines in the Post-Blakely Era." *Federal Sentencing Reporter* 17, no. 4 (2005): 233–42.
- Huq, Aziz. "Racial Equity in Algorithmic Criminal Justice." *Duke Law Journal* 68 (Forthcoming). <https://ssrn.com/abstract=3144831>.
- Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. "Human Decisions and Machine Predictions." NBER Working Paper No. 23180. Cambridge, MA: National Bureau of Economic Research, February 2017. <https://www.nber.org/papers/w23180>.
- Lance, Mark, and Margaret Little. "From Particularism to Defeasibility in Ethics." In *Challenging Moral Particularism*, edited by Mark Lance, Matjaž Potrc, and Vojko Strahovnik, 63–84. New York: Routledge, 2010.
- Laqueur, Hannah, and Ryan Copus. "Synthetic Crowdsourcing: A Machine-learning Approach to Inconsistency in Adjudication." Working paper, December 2017. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2694326.
- Leibovitch, Adi. "Punishing on a Curve." *Northwestern University Law Review* 111, no. 5 (2017): 1205–80.
- Little, Margaret Olivia. "Moral Generalities Revisited." In *Moral Particularism*, edited by Hooker Brad, and Margaret Olivia Little, 276–304. Oxford: Oxford University Press, 2000.
- Morris, Norval. *The Future of Imprisonment*. Chicago: University of Chicago Press, 1974.
- Rachlinski, Jeffrey J., Sheri Lynn Johnson, Andrew J. Wistrich, and Chris Guthrie. "Does Unconscious Racial Bias Affect Trial Judges?" *Notre Dame Law Review* 84, no. 3 (2009): 1195–1246.
- Reitz, Kevin R. "Comparing Sentencing Guidelines: Do US Systems Have Anything Worthwhile to Offer England and Wales?" In *Sentencing Guidelines: Exploring the English Model*, edited by Andrew Ashworth, and Julian V. Roberts, 182–201. Oxford: Oxford University Press, 2013.
- Roberts, Julian V. "Structuring Sentencing in Canada, England and Wales: A Tale of Two Jurisdictions." *Criminal Law Forum* 23, no. 4 (2012): 319–45.
- Roberts, Julian V., and Oren Gazal-Ayal. "Statutory Sentencing Reform in Israel: Exploring the Sentencing Law of 2012." *Israel Law Review* 46, no. 3 (2013): 455–79.
- Starr, Sonja B. "Evidence-based Sentencing and the Scientific Rationalization of Discrimination." *Stanford Law Review* 66, no. 4 (April 2014): 803–72.
- Tata, Cyrus. "The Application of Judicial Intelligence and 'Rules' to Systems Supporting Discretionary Judicial Decision-making." *Artificial Intelligence and Law* 6, no. 2–4 (1998): 203–30.
- Tata, Cyrus. "The Struggle for Sentencing Reform: Will the English Sentencing Guidelines Model Spread?" In *Sentencing Guidelines: Exploring the English Model*, edited by Andrew Ashworth, and Julian V. Roberts, 236–56. Oxford: Oxford University Press, 2013.
- Tonry, Michael. *Sentencing Matters*. Oxford: Oxford University Press, 1998.
- Yang, Crystal. "Free at Last? Judicial Discretion and Racial Disparities in Federal Sentencing." *Journal of Legal Studies* 44, no. 1 (2015): 75–111.