CrossMark

Do Theories of Punishment Necessarily Deliver a Binary System of Verdicts? An Exploratory Essay

Federico Picinali[1] (iD)

Abstract Scholars writing on theories of punishment generally try to answer two main questions: what human behaviour should be punished and why? Only cursorily do they concern themselves with the question as to how confident in the occurrence of criminal behaviour we must be prior to punishing—i.e., the question of the criminal standard of proof. Theories of punishment are ultimately theories about choices of action—in particular, about how to treat individuals. If this is correct, it seems that they should not overlook one of the fundamental variables governing human decision-making: the uncertainty about the facts relevant to our acting. Now, the question as to whether existing theories of punishment require a standard of proof as high as 'proof beyond a reasonable doubt' is gaining increasing attention in the scholarship. However, scholars working on theories of punishment give little attention to a particular way in which human decision-making handles the problem of uncertainty. In our everyday lives, we often decide in a many-valued, rather than a binary, fashion. Instead of having a single evidential threshold, the satisfaction of which determines whether we act or stay put, we tend to adjust our actions to our degree of confidence in certain states of affairs. In other words, we decide based on a ladder of evidential thresholds: the features of our actions vary according to the evidential threshold that we have satisfied. Notably, criminal trials do not follow this structure and theorists generally take this departure for granted. Why shouldn't trials work as '*ex post facto* bets,' whereby the response that the state is willing to 'wager' correlates with the fact finder's confidence in the defendant's guilt? The paper explores this question; in particular, it assesses whether the main theories of punishment (consequentialist, retributive, and communicative) necessarily deliver a binary system of verdicts. The work is part of a long-

✉ Federico Picinali
  f.picinali@lse.ac.uk

[1] Law Department, London School of Economics and Political Science, London, UK

🍀 Springer

term research project on the comparison between the binary and the many-valued models of the system of criminal verdicts.

Keywords Theories of punishment · System of verdicts · Binary decision-making · Many-valued decision-making · Criminal fact finding · Standard of proof

1 Introduction

Criminal trials in England and Wales—but the same holds true for several common law and civil law jurisdictions—are characterised by a simple 'binary' system of verdicts. There is a single, high evidential threshold—i.e., the 'reasonable doubt' or 'being sure of guilt' standard. If the threshold is met, the verdict is 'guilty' and the consequence of punishment is unleashed; if the threshold is not met, the verdict is 'not-guilty' and punishment does not ensue. The defendant who is found 'not-guilty'—or is not yet found 'guilty'—should be presumed innocent and treated accordingly. The single evidential threshold does not allow for a middle ground between a finding of 'not-guilty' and a finding of guilt: verdicts are categorical.

The situation is somewhat different in Scotland, where the law allows for an intermediate verdict of 'guilt not proved.'[1] Similarly, in the continental *jus commune* of the late middle ages and early modern era, criminal fact finding admitted of more than two outcomes, depending on the amount and the type of evidence available against the defendant. For instance, in the absence of 'full proof'—which consisted of two coherent testimonies or a confession—the judge could convict to a milder sentence—the so-called '*poena extraordinaria*'—if significant circumstantial evidence against the defendant was available.[2] The Scottish and *jus commune* systems of verdicts are 'many-valued.'[3] The essential feature of many-valued systems is that they consist of a ladder of two or more evidential thresholds, and corresponding ladders of verdicts and punishments. It is important to stress that in the many-valued model the ladder of verdicts and punishments does not track the presence of factors that mitigate or aggravate the defendant's responsibility; instead, it only tracks the presence of weaker or stronger evidence of responsibility, as represented by the evidential threshold that is satisfied each time.[4] The Scottish and *jus commune* systems, however, are only rough instantiations of the many-valued model: their evidential thresholds—and corresponding verdicts and punishments—are limited in number and not always clearly expressed or defined. I will soon present a more refined example of many-valued system.

---

[1] Cf. the old Italian Codice di Procedura Penale—enacted in 1930. Paragraph 3 of art. 479 authorised judges to deliver a verdict of 'acquittal for insufficiency of the evidence.' This was different from a 'full acquittal,' which was delivered when the hypothesis of innocence was more likely than that of guilt.

[2] For a brief treatment of this and other aspects of medieval continental criminal procedure see M. Meccarelli, 'Le Categorie Dottrinali della Procedura e l'Effettività della Giustizia Penale nel Tardo Medioevo,' in J. Chiffoleau, C. Gauvard, A. Zorzi (eds.), *Pratiques Sociales et Politiques Judiciaires dans les Villes de l'Occident à la fin du Moyen Âge* (Collection de l'Ècole Française de Rome, 2007).

[3] This term is used in logic to refer to systems admitting of more than two truth values. See S. Gottwald, 'Many-Valued Logic' (2015) Stan. Enc. Phil. 1. I thank John Davis for pointing out that the term that I was previously using—i.e., 'analogue'—was imprecise, as it referred to a continuum of values.

[4] Of course, the severity of punishment may well increase as the responsibility of the defendant increases, but this is not a definitional trait of the many-valued model.

This essay is part of a long-term research project consisting in a comparison between the binary and the many-valued models of the system of verdicts. This comparison will be undertaken both at the 'pragmatic' and at the 'theoretical' level. At the pragmatic level, my aim is to assess which model promises to fare better on a series of salient pragmatic parameters, which include, in particular, the ease of implementation and the prevention of crime. These two parameters refer, respectively, to a procedural and a substantive desideratum that are generally shared among criminal law theorists.[5] At the theoretical level, instead, my aim is to assess whether the two models presuppose a particular theory of punishment and political theory and—should one or both models be compatible with more theories—which theory of punishment and political theory best justifies each model. If these normative links are established, it will appear that the choice between the two models of the system of verdicts cannot be governed by 'neutral' pragmatic considerations only. It necessarily depends on our commitment to deeper views concerning the appropriate function of punishment and the values that this measure protects and fosters in the relationship between the citizen and the state. This essay is an initial contribution to the theoretical part of the research project. Here I explore the question as to whether a many-valued model of the system of verdicts may be compatible with theories belonging to the three main families of theories of punishment—consequentialist, retributive, and communicative. Cursorily, I also address the question as to whether such a model may better achieve the goals of these theories than a binary model.

There could be infinite variations on the many-valued model, depending on the number of thresholds characterising the system of verdicts, on how high these thresholds are, and on what are the consequences that may attach to each verdict. The binary model too may present infinite variations, depending on how high is the single threshold and on the consequences that may attach to each of the two verdicts. Here I am not particularly interested in fixing precisely any of these variables—although I accept that how they are fixed has ramifications on the question addressed in this paper. My focus is mainly on the meta-variable many-valued/binary, as defined by the distinctive essential features of each model, that I previously described. However, I appreciate that a discussion of the meta-variable is meaningless—if at all possible—without giving at least a rough description of the binary and many-valued systems that I have in mind for my investigation. As to the binary system, I will be referring to a system such as that of England and Wales. This system is characterised by the 'reasonable doubt' or 'being sure of guilt' standard—generally associated with a probability of guilt higher than .9—and inflicts punishments only for the guilty verdict—where punishments include measures such as community service, probation, fines, and imprisonment. As to the many-valued system, I will be referring to a system instantiating the many-valued features with greater accuracy and detail than the extant and historic cases mentioned above. For example, consider a many-valued system with four evidential thresholds, corresponding to the following probabilities of guilt: .5, .7, .8, and .9. If the .5 threshold is not passed, the verdict is 'innocent' and no punishment ensues. If the .9 threshold is passed, the verdict is 'guilty' and the punishments available for a 'guilty' verdict in the English and Welsh system are applicable. As for the thresholds .5, .7, and .8, the verdicts for passing them read 'there is a probability higher than—respectively—.5, .7, .8 that you are guilty.' The respective punishments are: the admissibility—with leave of the court—of the verdict as evidence of bad character in future proceedings for a similar offence; an exception to the double jeopardy guarantee

---

(e.g., the conditions for a new trial for the same crime to be possible would be less stringent than the current English and Welsh 'new and compelling evidence' test);[6] and an automatic sentence enhancement if the defendant is later convicted for another crime.

Here I am using numbers for the sake of simplicity. But words may be used instead. For instance, the .5 threshold identifies the state where innocence and guilt are 'equally likely.' Therefore, if this threshold is passed, guilt is 'more likely than' innocence. Also, in the case of a jury trial, the system may be designed in such way that the jury would not be required to apply any of these thresholds, and possibly would not even be told what the relevant thresholds are. The jury would simply be asked to express its assessment of the probability of guilt within a numerical or a verbal range. It would then be for the judge to apply the system of verdicts to the jury's finding. To be sure, these issues of design will be carefully addressed in the part of the research project concerning the implementation of the binary and the many-valued models of the system of verdicts.[7]

A quick terminological point is needed before moving forward. In the above, I used the term 'punishment' to refer to all afflictive measures inflicted by the system, whether they are associated with a categorical verdict of guilt or with an intermediate verdict. It is reasonable to argue, though, that we should call 'punishment' only the afflictive measures associated with the categorical guilty verdict. The reason for this lies in the concept of 'punishment.' Among the essential elements of this concept is that according to which punishment is issued *because of what the defendant did*. In other words, punishment is *for a deed*.[8] A corollary of this element is that, as a conceptual matter, we cannot *punish* the defendant, e.g., 'for having probably committed a crime,' or 'because there is a probability of .5 that she is guilty.' We can only *punish* the defendant 'for having committed the crime,' without adding any qualification.[9] True, we cannot be certain that the crime occurred; in fact, any practicable standard of proof falls below certainty. However, if we want to see ourselves as *punishing* an individual, we must behave as if we were certain.[10]

---

[6] Cf. s. 78 Criminal Justice Act 2003.

[7] An interesting alternative to the implementation of a many-valued system would be to create one or more lesser offences corresponding to each crime, such that the new offences consist in only some of the elements of the crime. These offences would be punished more lightly than the complete crime, so as to reflect the remaining uncertainty as to whether those who have committed the former are also guilty of the latter. Under this alternative, the substantive criminal law would be used to create a regime that is similar to the many-valued model of the system of verdicts in that it adjusts punishment according to the probability of guilt. A theory of this kind is developed in D. Teichman, 'Convicting with Reasonable Doubt: An Evidentiary Theory of Criminal Law' (2017, available at SSRN: https://ssrn.com/abstract=2932743)—Teichman's goal, though, seems chiefly descriptive. In fact, such an alternative theory would raise significant problems— which I can only mention briefly here. First, under this alternative, substantive criminal law would be unwieldy and, probably, messy—absent exceptional and, frankly, unrealistic rigour on the part of the law-maker. Second, the prohibition of much (apparently) harmless behaviour—through the lesser included offences—would confuse the citizens as to what values the criminal justice system protects. Third, cases would occur where, notwithstanding that it is evident that the final crime has not been committed, the lesser—possibly non-blameworthy—offence has. In these cases, it is unclear whether the theory would justify convicting the defendant. Finally, an alternative of this sort does not seem to allow for the same flexibility in design that characterises the many-valued model of the system of verdicts.

[8] Cf. H.L.A. Hart, *Punishment and Responsibility: Essays in the Philosophy of Law* (Oxford University Press, 2008) 4–6, and David Boonin, *The Problem of Punishment* (Cambridge University Press, 2008) 17–21, addressing, and rejecting, at 20–21 the possible consequentialist objection according to which punishment should not be defined so as to comprise this element, otherwise consequentialist theories of punishment—for which the link between punishment and deed is not essential—are ruled out as a matter of definition rather than being assessed for their qualities.

[9] See *id*, 19–20, suggesting that punishment requires that the defendant *be believed* to have broken the law.

[10] See *ibidem*.

We must 'make a leap' from a qualified probabilistic conclusion—corresponding to the criminal standard of proof—to a categorical conclusion.[11] A categorical verdict of guilt seems, therefore, necessary for punishment. Notwithstanding this, for the sake of simplicity, I will continue using the term 'punishment' also to refer to afflictive measures associated with intermediate verdicts. This usage is imprecise, but as long as the reader is aware of the terminological point just made it should produce benefits at no cost.

2 Why Does the Project Matter?

The aim of the comparison between the binary and the many-valued models of the system of verdicts is ultimately normative. I intend to answer the question as to whether we should switch to a many-valued model. But why should we bother with this comparison in the first place?[12] Is the current binary system giving us reasons to worry? Or is this just the whimsical research project of a bored academic? What are we to gain from it?

The first answer to this set of questions is that the comparison will at the least provide a *pro tanto* justification for the current state of affairs. If the many-valued model turns out to be an unpalatable option, our current adoption of a binary model gains support. Does it need support?—someone may ask. Interestingly, aside from a few exceptions,[13] most scholars seem to take the binary model of the system of verdicts for granted, as if the reasons for having it were self-evident. I don't think that they are.[14] Work needs to be done

---

[11] For an argument that such a leap is justified only if proof of guilt is beyond a reasonable doubt, see F. Picinali, 'Is the Reasonable Doubt Standard Justified? A Reconstructed Dialogue' (2017, unpublished manuscript on file with the author).

[12] Vincent Chiao raised the question as to whether a many-valued system of verdicts already exists. More precisely, if we look beyond the criminal justice system, and consider also the civil, the administrative, and the disciplinary systems, we may realise that a system with different evidential thresholds, verdicts, and punishments is already in place: it consists precisely in the intersection of these systems. However, this intersection is not functionally equivalent to a criminal many-valued system of verdicts. First, the overlap between the different normative systems is incomplete—i.e., not every crime is also a tort or an administrative or disciplinary violation; and, even when there is an overlap in the law, this may not be realised in practice—mainly because the decisions as to whether to start the different proceedings are the prerogatives of different authorities/individuals. Thus, the intersection of different normative systems is not co-extensive with a criminal many-valued system of verdicts. Second, it is reasonable to argue that the functions of criminal law, of tort law, of administrative law, and of disciplinary law are different. Thus, the intersection of normative systems and proceedings could not possibly perform the same role—and achieve the same results—as a criminal many-valued system of verdicts.

[13] See A.D. Leipold, 'The Problem of the Innocent, Acquitted Defendant' (2000) 94 Nw. U. L. Rev. 1297, S. Bray, 'Not Proven: Introducing a Third Verdict' (2005) 72 U. Chi. L. Rev. 1299, H. Lando, 'The Size of the Sanction Should Depend on the Weight of the Evidence' (2005) 1 Rev. Law & Econ. 277, L. Laudan, 'Need Verdicts Come in Pairs?' (2010) 14 E & P 1, T. Fisher, 'Constitutionalism and the Criminal Law: Rethinking Criminal Trial Bifurcation' (2011) 61 U. To. L. J. 811, and T. Fisher, 'Conviction without Conviction' (2011) 96 Minn. L. Rev. 833. See also A.J. Kolber, 'Smooth and Bumpy Laws' (2014) 102 Cal. L. Rev. 655, in particular at 678–680.

[14] Claiming that the presumption of innocence is such a self-evident reason would be question-begging. The presumption of innocence—at least as it is formulated in art. 6(2) ECHR—does not prescribe any model of the system of verdicts, or any standard of proof for that matter. In order to defend the binary model via the presumption, we must first give an account of 'innocence,' of when it should be protected, and of how it should be protected. This account requires answering questions such as whether the intermediate verdicts and punishments that I previously described are compatible with 'treating someone as innocent until proven guilty'; and whether they are more detrimental measures than, e.g., depriving a suspect of her liberty during the investigation—which is generally considered compatible with the presumption. Moreover—although I accept that this may be viewed as a bold claim—were we to conclude that the presumption of innocence

in order to ensure that one of the crucial features of our criminal justice system is justified—that this feature is not merely the product of distant historical developments and of inertia.

The issue of the justification of the binary model becomes even more interesting if we consider that in our everyday deliberations we often prefer many-valued decision-making: rather than making black-or-white decisions, we adjust our actions so as to reflect the confidence that we have in the truth of some states of affairs. For instance, if we plan to cycle to work and there is an appreciable probability of rain, it is unlikely that our agency will be the result of a decision between two options exclusively (i.e., cycling and not cycling), where the choice between the two depends on the satisfaction of a single probability threshold. Instead, we are more likely to adjust our action in light of the probability of rain: the higher is the probability, the more and the heavier are the waterproof garments that we are willing to bring on the ride. True, we may decide not to cycle if the probability of rain is very high, but not cycling would be just one of the several options available.[15] What features of criminal trials speak against adopting this decision-making strategy? Why shouldn't trials work as '*ex post facto* bets,' whereby the response that the state is willing to 'wager' correlates with the fact finder's confidence in the defendant's guilt?

The second answer to the set of critical questions with which I started this section draws on the consideration of some problems affecting the current binary system of verdicts. Apparently, these problems would not affect—or would be less serious in—a many-valued system such as the one previously described. This fact is hardly a sufficient reason to switch to such a system. However, it is certainly an indication that the comparison should be taken seriously. Indeed, the comparison may ultimately suggest that we abandon the binary model or it may point our attention to ways in which these problems can be alleviated, if not eliminated, without such a radical change in the law. The problems to which I am referring include the following:

(a)   Insufficient Exoneration[16]

In a binary system with a high standard of proof, true negatives are not sufficiently exonerating. Given the little information that they provide, 'not-guilty' verdicts are susceptible of different readings on the part of the public and institutions alike: in particular, one may believe that the defendant is innocent or that she is guilty and that the evidence

---

Footnote 14 continued

mandates the binary model, we may still decide to dispense with the presumption if the balance of reasons is in favour of switching to the many-valued model.

[15] Neha Jain pointed out that the bike ride example differs from the case of the criminal trial in at least one important respect: the former is an instance of forward-looking fact finding, whereas in the latter fact finding is backward-looking. It is possible that the forward-looking nature of fact finding incentivises many-valued decision-making. However, I do not think that the mere fact that fact finding is backward-looking justifies the adoption of binary decision-making. Indeed, there are examples of many-valued backward-looking fact finding. Consider the case of a mountaineer who does not know whether it snowed on a particular slope that she intends to climb: the kind and amount of gear that she decides to bring for the ascent may depend on how confident she is that the slope is covered in snow. In any case, the relationship between the variables 'forward-/backward-looking' and 'many-valued/binary' needs a careful analysis, which I cannot undertake here.

[16] On this problem cf. Leipold, *supra* note 13, Bray, *supra* note 13, at 1320–1326, and Laudan, *supra* note 13, at 2–10. In the US, the problem discussed here is exacerbated by the presence of a gap between the public understanding of the reasonable doubt standard and the way in which the standard is understood and applied by jurors—the standard being less stringent according to the latter understanding. See A. Walen, 'Proof Beyond a Reasonable Doubt: A Balanced Retributive Account' (2015) 76 Louisiana L. Rev. 355, at 373–376.

produced at trial was not sufficient to meet the high standard of proof. Notice that the informational deficiencies of the 'not-guilty' verdict do not depend on how it is phrased, but on the high evidential threshold discriminating between the two verdicts. Because of the possible alternative readings of a 'not-guilty' verdict, the innocent defendant who has been acquitted may still bear negative consequences in terms of spoiled reputation, damaged social or family ties, professional hurdles, etc. For instance, in several US jurisdictions, acquitted defendants can be legally denied credit by credit agencies, can be disqualified from adopting children, and can be obliged to report their prior arrest to prospective employers, if asked to do so.[17] Under a system with one or more intermediate verdicts and evidential thresholds, the situation would be different. In a many-valued system, as the one described in the Introduction, the 'innocent' verdict issued when the .5 threshold has not been met would clearly signal that the defendant is unlikely to be responsible for the crime, thus barring the aforementioned negative consequences from taking place.[18]

(b)   The Waste of Resources

There is an additional problem concerning the informativeness of binary verdicts. As already mentioned in (a), the 'not-guilty' verdict provides little information. More precisely, the verdict does not accurately reflect the fact finder's assessment of the evidence. This information would convey the likelihood of guilt and may be useful to fellow citizens and to institutions when they make decisions concerning the defendant after the trial[19]—e.g., whether to befriend the defendant, whether to continue or start a job partnership, whether to grant benefits, whether to allow the defendant to adopt a child. The problem is not just that a binary system does not communicate sufficient information on the likelihood of guilt. It is also that this information is actually available, consisting in the basis and, more importantly, in the result of the jury's assessment. In not communicating it, the system wastes important resources. A many-valued system that is sufficiently fine-grained would not incur such a waste: it would make the information available through the ladder of verdicts. A possible alternative solution to this problem would be to adopt reasoned verdicts. I am not going to assess this proposal here. It suffices to say that, while it may address the problem of information waste (provided that the reasons given are sufficiently intelligible), adopting reasoned verdicts may not adequately solve the problem in (a). Nor would it solve the problems in (c) and in (d), to which I now turn.

(c)   The Aggregate Significance of Fact Finding Mistakes

A many-valued system may produce mistakes that would not result from a binary system. In particular, whilst in a binary system an innocent defendant who the fact finder considers very likely to be guilty—although not guilty beyond a reasonable doubt—would simply be acquitted, in a many-valued system she may be punished according to an intermediate verdict. Indeed, strictly speaking all the intermediate verdicts would be mistakes, given that a crime either did or did not occur—it did not '.6 or .7 occur.' This may very well mean that a many-valued system produces more mistakes than a binary

---

[17] See Laudan, *supra* note 13, at 7 and Leipold, *supra* note 13, at 1304–1313.

[18] I acknowledge that the mere fact of being brought to trial produces negative consequences for the innocent defendant, especially in terms of her reputation and social life. However, a many-valued system of verdicts as the one described promises to do a better job at neutralising (at least some of) these consequences than the current binary system, given that it allows for more effective exoneration.

[19] Cf. Bray, *supra* note 13, at 1307–1314.

system. However, the presence of a ladder of verdicts and of corresponding evidential thresholds and punishments should be expected to decrease the aggregate significance of the mistakes produced by the system. Consider the many-valued system described in the Introduction. The innocent defendants who are acquitted would bear no consequence. The innocent defendants who are punished according to intermediate verdicts, would only bear the moderate consequences attaching to these. Moreover, in the run of cases we should expect these outcomes to be rare, given that the intermediate thresholds correspond to probabilities of guilt that are as high as, or higher than, .5. On the other hand, it is unlikely that the guilty would be acquitted—an outcome that one would expect to be quite frequent under the current binary system. The large majority of guilty defendants would be either found guilty or would receive an intermediate verdict with the corresponding punishment. In light of these considerations, it is reasonable to argue that the additional harm that the many-valued system at issue would produce by punishing the innocent through interme-diate verdicts would be outweighed by the system's ability to reduce the harm consisting in the acquittal of the guilty. As I will try to show in the next section, this conclusion holds true irrespective of which theory of punishment we resort to in order to characterise the 'harm' consisting in the false positive and in the false negative. The success of this argument, though, seems to depend on the truth of the following assumptions: (1) that the majority of defendants who go to trial are guilty; and (2) that the evidence that is presented in trials of guilty defendants is generally stronger (and thus indicates a higher probability of guilt) than the evidence that is presented in trials of innocent defendants. If these assumptions were false—e.g., if 50% of defendants going to trial were innocent and if the evidence presented at their trials were generally as strong as the evidence presented at the trials of the guilty—the harm that the many-valued system would avoid through convicting more guilty defendants would be balanced out, if not outweighed, by the additional harm that the system would produce through convicting more innocent defendants. The many-valued system considered here would not, therefore, reduce the aggregate significance of fact finding mistakes. Being unaware of reliable evidence on the accuracy of these assumptions—if such evidence can be obtained at all—I can only recognise that they are both reasonable assumptions to make.[20] Moreover, if these assumptions were false, the justification of our criminal justice system would be undermined by features of it that are possibly more fundamental than its system of verdicts—e.g., inadequate arrest and charging laws and practices, and the ineffective operation of the investigative process. True, under such a faulty system adopting the many-valued model would probably increase the aggregate significance of fact finding mistakes. But the system would be already unjustified for independent reasons, and fixing it would require much more incisive reforms than a switch to a different model of the system of verdicts. I accept that in such dire circumstances discussing this switch should not be a priority. I hope, though, that these circumstances do not obtain in our criminal justice system, and I proceed on this basis.

---

[20] A member of the audience at the Center for Transnational Legal Studies objected that in a criminal justice system like the US, characterised by a high rate of guilty pleas, the majority of guilty people probably plead guilty, whereas the majority of innocent people probably do not; and that, therefore, we should expect that those who go to trial are for the most part innocent. This argument, though, is itself based on unver-ifiable claims. Moreover, it is a *non sequitur*. The premises do not include information about the proportions of, respectively, guilty and innocent people who are charged with crimes in the first place. If guilty people are the overwhelming majority of those who are charged, the assumption I make about the population of those who go to trial may still hold even if most guilty people plead guilty and most innocent people do not. An aside: the rate of guilty pleas may decrease with a many-valued system, since some of those who would plead guilty in the current binary system may prefer attempting to gain an intermediate verdict to pleading guilty.

(d)   The Distortion of Fact Finding

Although theoretically simple, the binary model poses a serious practical issue. The fact finder is presented with a stark alternative between guilt/punishment, on the one hand, and non-guilt/non-punishment, on the other. The outcome depends on whether the single evidential threshold is met. The presence of an all-or-nothing alternative may put the fact finder under heavy pressure when—quite irrespective of the evidence produced at trial— she considers it desirable to punish or incapacitate the defendant. This may bring the fact finder to factor practical reasons into her theoretical decision—e.g., the desire to punish a dangerous defendant may be treated as a reason to find the defendant guilty, when the evidence of guilt is not sufficient to satisfy the reasonable doubt standard.[21] This reasoning produces an obvious distortion of fact finding, given that a desire to achieve a particular result is certainly not a valid reason to conclude that the defendant committed a particular act. Less pressure would be produced in a many-valued system, because its ladders of verdicts and of punishments make it possible to punish the defendant even when the evidential threshold that has been satisfied is lower than proof beyond a reasonable doubt. Thus, a many-valued system—when properly crafted—would be less likely to produce this distortion of fact finding.[22]

At a first glance, it seems that one would acknowledge these problems of the current binary system irrespective of the theory of punishment that she endorses. Other things being equal, consequentialists, retributivists, and supporters of communicative theories would probably prefer a system producing more effective exoneration of the innocent, less waste of resources, a less significant aggregate of mistakes, and fewer fact finding distortions. This system may be precisely some version of the many-valued model of the system of verdicts. This conclusion, however, is tentative and imprecise. In fact, what I said so far is not even sufficient to establish that a many-valued model would be compatible with theories belonging to each of the three families. This question requires a more careful assessment.

3 Theories of Punishment and the Many-Valued Model

Scholars writing on theories of punishment generally try to answer two main questions: *what* human behaviour should be punished and *why*? Only cursorily do they concern themselves with the question as to how confident in the occurrence of criminal behaviour we must be prior to punishing—i.e., the question of the criminal standard of proof. Theories of punishment are ultimately theories about choices of action—in particular, about how to treat individuals. It seems, therefore, that they should not overlook one of the fundamental variables governing human decision-making: the uncertainty about the facts relevant to our acting—in particular, whether the defendant did it or not. Now, the question

---

[21] In fact, practical reasons may influence fact finding in various ways. For instance, they may bring the fact finder consciously to disregard her task of ascertaining facts; they may bring the fact finder to give more weight to some evidence of guilt than this evidence should be given; they may bring the fact finder to interpret the standard of proof as if it were lower than it actually is.

[22] This is not to deny that a many-valued system may cause other forms of distortion. For instance, it may trigger particular biases on the part of the fact finder—e.g., a bias in favour of compromising, by choosing an intermediate verdict. For a discussion of the negative and positive effects of these biases, see Bray, *supra* note 13, at 1314–1320 and Fisher, 'Constitutionalism and the Criminal Law,' *supra* note 13, at 833–838. These and others problems concerning the implementation of the many-valued model will have to be considered during the pragmatic part of the research project.

as to whether existing theories of punishment require or justify a standard of proof as high as 'proof beyond a reasonable doubt' is gaining increasing attention in the scholarship.[23] However, there is very little research addressing the question as to whether theories of punishment are compatible with a many-valued model of the system of verdicts.[24] This is particularly curious if one considers that—as pointed out earlier—in our everyday decision-making we often deal with the problem of uncertainty in a many-valued fashion. In this section, I address this compatibility issue. I will try to show that the many-valued option does not encounter in-principle objections from any of the main families of theories of punishment.

3.1 Consequentialism

A purely consequentialist theory of punishment holds that the justification of punishment depends exclusively on the consequences that punishment produces. More precisely, punishment is justified if its consequences maximise the good. A reasonable conception of the good—and one that is generally supported by consequentialist theorists on punishment—would value, especially, the exoneration of the innocent and the prevention of crime. Can it be shown uncontrovertibly that the binary model—and, in particular, the current binary system—maximises these valuable consequences? This is yet to be demonstrated, and there are reasons to suppose that a many-valued system would better serve these goals.[25]

It is reasonable to expect that under a many-valued system similar to the one previously sketched more innocent defendants would be effectively exonerated than are under the current system. Indeed, it is reasonable to expect that many innocent defendants would be able to make strong enough cases to bring the probability of guilt below the .5 threshold and, therefore, would receive 'innocent' verdicts. On the other hand, many guilty defendants who would be acquitted under the current binary system, would in that many-valued system be punished according to an intermediate verdict. This should increase deterrence and, ultimately, crime prevention. Of course, the consequentialist may depart from my token of many-valued system and craft the ladder of evidential thresholds, verdicts, and punishments that achieves stigmatisation, deterrence, and incapacitation most effectively. Provided that the system is carefully designed, it is reasonable to expect that these ends would be more easily secured than in the current binary system.[26] Indeed, if properly harnessed, the many-valued model makes it easier to punish the guilty. The stigmatisation

---

[23] Consider, in particular, J. Reiman and E. van den Haag, 'On the Common Saying that it is Better that Ten Guilty Persons Escape than that One Innocent Suffer: *Pro* and *Con*' (1990) 7 Social Philosophy & Policy 226, L. Laudan, 'The Rules of Trial, Political Morality, and the Costs of Error: Or, Is Proof Beyond a Reasonable Doubt Doing More Harm than Good?' in L. Green and B. Leiter (eds.), *Oxford Studies in Philosophy of Law* (Oxford University Press, 2011), and Walen, *supra* note 16.

[24] As far as I am aware, the only work that addresses this question is Fisher, 'Conviction without Conviction,' *supra* note 13. Some criticisms of this work are advanced in the following footnotes.

[25] True, a reasonable conception of the good would value also other consequences, and these may be maximised by a binary system. This would raise the question as to what consequences should be prioritised. I leave this discussion for another time.

[26] For an argument that a properly crafted many-valued system would increase deterrence, see Fisher, 'Constitutionalism and the Criminal Law,' *supra* note 13. Fisher builds her argument on Lando, *supra* note 13—arguing in favour of variable sanctions, but only above the threshold of proof beyond a reasonable doubt. A problem with Fisher's and Lando's theories is that they rest on debatable assumptions concerning the information available to potential offenders, and their decision-making process, e.g., the assumption that knowledge of the actual guilt or innocence of defendants is available to potential offenders and the

of the guilty and—especially—deterrence and incapacitation should contribute to the prevention of crime.

Now, if the exoneration of the innocent and the prevention of crime are constitutive of the good, other things being equal the consequentialist should prefer the model of the system of verdicts that is more efficient in achieving these goals, that is—I suggest—the many-valued model. In fact, other things would not be equal: through the intermediate verdicts a many-valued system stigmatises and punishes innocent people who would be acquitted under the current binary system. However, if the (reasonable) assumptions mentioned in Sect. 2(c) are accurate, we should expect this additional evil to be outweighed by the additional good produced by the many-valued system in terms of more effective exoneration of the innocent and of increased crime prevention.

The appeal that the many-valued model may have for the consequentialist seems even more evident if we consider the decision-theoretic approach, that is, a consequentialist approach based on the calculus of expected utilities. While the discussion so far has concerned the maximisation of the good in the run of cases dealt with by the criminal justice system, decision-theory allows us to take a different perspective: that concerning the maximisation of the good in the individual case. Many would agree that punishing the guilty is at least *prima facie* good, while punishing the innocent is at least *prima facie* evil.[27] However, we are never certain about the defendant's guilt—and, therefore, her innocence. Thus, when inflicting punishment, we cannot be certain as to whether we are bringing about good or evil. If our goal is to maximise the good, it is not sufficient for us to determine which consequences of our actions constitute such maximisation; before making a choice of action, we also need to assess the likelihood of these consequences. Under the decision-theoretic approach, our choice of action should maximise the 'expected utility,' that is, the product of the value of each possible consequence and of its likelihood. When an action may have more than one consequence—which seems always to be the case—the expected utility associated with the action is the sum of the expected utilities corresponding to each consequence. Consider the case of the criminal trial. The action of convicting may produce the conviction of the guilty or the conviction of the innocent. Thus, the expected utility associated with this action is the sum of the expected utilities corresponding to each of these two possible consequences. The same reasoning applies to acquittals. If the probability of guilt is very high, convicting the defendant yields a high expected utility[28]: it is very likely that this action will produce a beneficial consequence (i.e., convicting the guilty) and very unlikely that it will produce a detrimental consequence (i.e., convicting the innocent). On the other hand, acquitting the defendant under these circumstances yields a low expected utility[29]: it is very likely that it will produce a detrimental consequence (i.e., acquitting the guilty) and very unlikely that it will produce a beneficial consequence (i.e., acquitting the innocent).

According to a decision-theoretic justification of the reasonable doubt standard, this standard represents the probability of guilt at which the expected utility of acquitting is the

---

Footnote 26 continued

assumption that criminality results from rational risk-assessment. To be fair, the latter assumption is typical of deterrence theories.

[27] I say '*prima facie*' because there may be ulterior consequences that counteract, respectively, the good and the evil of these outcomes.

[28] Or, at least, a relatively high one, if compared to the expected utility of acquitting.

[29] Or, at least, a relatively low one, if compared to the expected utility of convicting.

same as the expected utility of convicting.[30] Only if the probability of guilt is the same as, or is higher than, the reasonable doubt standard, conviction is warranted, because only in these circumstances conviction maximises expected utility.[31] Now consider a binary system with the reasonable doubt standard. Let's assume—according to the consequentialist justification of the standard that I just mentioned—that if the probability of guilt is below the reasonable doubt standard the action that maximises the expected utility is acquitting the defendant. After all, should we convict the defendant and should the defendant be innocent—a hypothesis that is reasonable under the circumstances—we would be inflicting severe punishment on her—e.g., imprisonment, or a fine, or both. But what if we could inflict less serious punishment when the probability of guilt is just below the reasonable doubt standard? What if we devised a ladder of punishments, each corresponding to a different probability threshold, such that the lower is the threshold the less severe is the punishment? In other words, what if we adopted a many-valued system of verdicts? If the ladder of punishments and of thresholds is well calibrated, it may be that, when the probability of guilt is that indicated by one of the thresholds below proof beyond a reasonable doubt, inflicting the corresponding punishment would yield a higher expected utility than acquitting—as the current binary system would require.[32] To refer to my sketch of many-valued system, when the probability of guilt is above .8 and equal to, or below, .9, inflicting on the defendant the prospect of an automatic sentence enhancement for the case that she is later convicted for another crime may yield a higher expected utility than simply letting the defendant walk away. After all, it is very likely that the defendant is guilty—and thus the appropriate recipient of punishment—and punishment would afford a quantum of stigmatisation and deterrence. In the unlikely case that the defendant were innocent, the kind of punishment that she would receive may not have much impact on her, and may not be decisively different from the lack of exoneration characterising a not-guilty verdict under the current binary system.

Of course, the calculus of expected utilities is no easy task. Producing a ladder of punishments and of thresholds that maximises expected utility would be a taxing job for the consequentialist—if at all possible. However, on an abstract level the many-valued model seems to have great appeal for the consequentialist. From a decision-theoretic perspective it is hard to make sense of the single cut-off point and of the black-or-white alternative characterising the binary model. At least when guilt is more likely than not, expected utility seems maximised by accompanying every increase in the probability of guilt with an increasing quantum of punishment. In other words, maximising expected utility seems to demand that, as the probability of inflicting punishment on an innocent individual decreases, the seriousness of punishment increases up to the point where punishment takes the form that seems appropriate for the case when guilt is proven beyond a reasonable doubt.

---

[30] For a criticism of this claim by a consequentialist, see Laudan, *supra* note 23.

[31] See F. Picinali, 'Two Meanings of Reasonableness: Dispelling the "Floating" Reasonable Doubt' (2013) 76 Modern Law Review, 846–847. On the application of decision theory to criminal fact finding, see also J. Kaplan, 'Decision Theory and the Factfinding Process' (1968) 20 Stanford Law Review 1065, L. H. Tribe, 'Trial by Mathematics: Precision and Ritual in the Legal Process' (1971) 84 Harvard Law Review 1329, 1378 ff., E. Lillquist, 'Recasting Reasonable Doubt: Decision Theory and the Virtues of Variability' (2002) 36 UC Davis Law Review 85, and L. Laudan and H.D. Saunders, 'Rethinking the Criminal Standard of Proof: Seeking Consensus about the Utilities of Trial Outcomes' (2009) 7 International Commentary on Evidence 1.

[32] Cf. Lando, *supra* note 13, at 283–284.

3.2 Retributivism

At a first glance, it seems that any retributivist would object to the many-valued model of the system of verdicts. Any retributivist would subscribe at least to the claim that desert is a necessary condition for punishment. She would, therefore, object to a model—such as the many-valued—that seemingly countenances the conviction of the innocent, as it sets a ladder of intermediate verdicts and punishments. And yet, also a binary system with a standard as high as the reasonable doubt standard countenances false convictions. Not only are false convictions envisaged as a possibility—after all, the law does not require certainty of guilt for a conviction—but we can be confident that in the long run innocent people will be convicted even with such a demanding standard of proof—surely the recent experience of DNA exonerations in the US reinforces this expectation. The awful truth is that no workable criminal justice system can avoid the conviction of the innocent. Therefore, the retributivist finds herself in the unsavoury position of having to choose between, on the one hand, accepting the risk of convicting the innocent for the sake of convicting the guilty, and thus relinquishing her claim that desert is a necessary condition for punishment; and, on the other hand, sticking to this claim but giving up with the enterprise of punishment all together.

A retributivist may deny this problem—sometimes dubbed the 'fallibility argument'—by claiming that only the intentional or knowing conviction of an identified innocent person would run afoul of the retributivist commitment not to punish the innocent.[33] Instead—this retributivist would claim—it is acceptable to set up a criminal justice system even if we know that some statistical lives will be falsely convicted if we do so.[34] This position may be tenable, but it does not fully take up the challenge of the fallibility argument: what is for the retributivist an acceptable degree of risk of false convictions and how should this degree be calculated?[35]

Some retributivists acknowledge and take up the challenge. Besides accepting that the innocent may be convicted—possibly also intentionally[36]—they recognise that the question as to what is the acceptable degree of risk of punishing innocent people should—or at least may—be part of the retributivist agenda. Thus, Larry Alexander theorises as a viable retributivist account the account according to which negative and positive desert can be compared, and we should devise a system of punishment that 'comes closest to placing people in their proper comparative positions.'[37] Under this view—rather than withholding punishment because of the risk of punishing the innocent—we may accept some instances

---

[33] This in fact represents a departure—or a redrafting—of the principle according to which desert is a necessary condition for punishment.

[34] See M.S. Moore, 'Justifying Retributivism' (1993) 27 Israel Law Review 15, at 20. For a critical assessment of this claim, see D. Dolinko, 'Retributivism, Consequentialism, and the Intrinsic Goodness of Punishment' (1997) 16 Law and Philosophy 507.

[35] Moore accepts that a retributivist may be a consequentialist. Provided that she does not violate the commitment not to intentionally or knowingly punish identified innocent people, she may accept a trade-off between trial outcomes so as to maximise the intrinsic good of giving people what they deserve. However, Moore concludes these reflections by saying that '[w]here in the actual design of punishment institutions the consequentialist-retributivist comes out on this balance need not here detain us' (Moore, *supra* note 34, at 19). In a footnote to this passage, he gives brief and tentative indications as to how a retributivist may approach the calculus.

[36] See D. Husak, 'Retributivism *In Extremis*' (2013) 32 Law and Philosophy 3, at 15–16.

[37] L. Alexander, 'Retributivism and the Inadvertent Punishment of the Innocent' (1983) 2 Law and Philosophy 233, at 237.

of undeserved punishment in order to bring about a state of affairs that comes closest to giving everyone what she deserves. Jeffrey Reiman claims that retributivists 'must … compare the relative size of evils when they must choose between them' and that 'retributivism goes awry when it refuses to allow such relative measures.'[38] In other words, decisions such as whether to set up a criminal justice system, whether to adopt a high standard of proof, or whether to punish an identified person should all depend on the calculation of, and the comparison between, the possible evils that these actions may bring about—in particular, not giving people what they deserve and giving people what they do not deserve.

While these retributivist accounts have the merit of acknowledging the realities of punishment—i.e., every workable system of punishment will produce false positives—they inevitably raise the question of the relationship between consequentialism and retributivism. Aren't these accounts consequentialist? After all, they recommend that choices of action be made in light of the consequences that such choices may bring about. In particular, they recommend that a system of punishment is set up so as to maximise a particular good—i.e., giving everyone what she deserves.[39] Why, then, shouldn't a retributivist subscribe to a decision-theoretic model such as the one sketched in the previous section? In fact, it seems that a decision-theoretic model is viable under retributivism; however, there are several alternative grounds based on which a retributivist decision-theoretic model may be distinguished from a consequentialist one. A retributivist may contend that the only consequences that should be taken into account in the calculus of the acceptable degree of risk of false convictions are the immediate—or intrinsic—consequences of punishment, consisting in giving people what they do or do not deserve. The ulterior consequences—e.g., deterrence or the suffering caused to people other than the defendant—are, instead, irrelevant—while they would be considered under a consequentialist account.[40] Alternatively, a retributivist may recognise that ulterior consequences are indeed reasons for or against punishing. However, she may claim that ulterior consequences speaking *in favour of* punishing cannot be factored into the calculus of the acceptable degree of risk of convicting the innocent.[41] This is because these consequences can never count as reasons to punish the innocent: only if desert has been established can they count as reasons to punish.[42] Finally—but this has no pretence of being an exhaustive list—a retributivist may accept that any ulterior consequence is included in the calculus, but recognise some kind of lexical priority to the intrinsic consequences.[43]

---

[38] Reiman and van den Haag, *supra* note 23, at 231.

[39] Cf. J. Gardner, 'Introduction' of Hart, *supra* note 8, at xvi.

[40] Cf. Reiman and van den Haag, *supra* note 23, at 231–232.

[41] Instead, this retributivist may factor into the calculus the ulterior consequences speaking against punishing.

[42] See Walen, *supra* note 16, at 426–430. Cf. Alexander and Ferzan's notion of 'moderate retributivism' in L. Alexander, K. Ferzan, *Crime and Culpability* (Cambridge University Press, 2009), at 7–8 and Husak's 'Why Punish the Deserving?' in D. Husak, *The Philosophy of Criminal Law: Selected Essays* (Oxford University Press, 2010), in particular at 394–399. According to these views, desert is a necessary and sufficient condition for punishment, but it does not mandate punishment. Once desert is established, consequentialist considerations will determine whether punishment should be inflicted.

[43] Cf. M.N. Berman, 'Punishment and Justification' (2008) 118 Ethics 258, in particular at 285–286—arguing that the good that is intrinsic in punishing the guilty cancels out (rather than merely overriding) all the ulterior consequences speaking against punishing. See also Berman's account of 'retributive instrumentalism' in M.N. Berman, 'Two Kinds of Retributivism' in R.A. Duff and S. Green, *Philosophical Foundations of Criminal Law* (Oxford University Press, 2011).

Where does this discussion leave us with respect to the question as to whether retributivism is compatible with a many-valued model of the system of verdicts? So far, we have seen that retributivists may accept the risk of punishing the innocent, and that they may identify the acceptable degree of risk by weighing the good and the bad consequences of punishing or failing to punish—whichever of the three abovementioned alternatives a retributivist may follow. If we add to this picture the typical retributive corollary that punishment should be proportionate to the individual's desert, we should begin to see why a many-valued model might be an appealing option for the retributivist. The higher the probability of the defendant's guilt, the more likely we are to incur the evil of not giving her what she deserves should we decide to acquit. If—starting at least from a probability of guilt higher than .5,[44] as in the token of many-valued system presented earlier—we inflicted on the defendant a quantum of punishment that increases as the probability of guilt increases, we would be reducing the overall magnitude of the intrinsic evil of false acquittals that we would incur under a binary system with a high standard of proof. More guilty defendants would get at least part of what they deserve: as the probability of guilt increases, their punishment would be progressively closer to a punishment that is proportionate to their crime.[45] True, under a many-valued system we would incur the evil of punishing more innocent people. These occurrences, however, would still be limited if the probability thresholds corresponding to the intermediate verdicts are sufficiently high (and thus indicate that the defendant is probably guilty). Also, one should consider that the punishments corresponding to the intermediate verdicts would be substantially less severe than the punishments triggered by proof beyond a reasonable doubt, and that their severity would decrease as the probability of guilt decreases. In light of these considerations, it is plausible to argue that the additional evil that a many-valued system produces by punishing innocent people through intermediate verdicts would be outweighed by the system's ability to reduce the evil consisting in the acquittal of the guilty[46]—the success of this argument, though, depends on whether the (reasonable) assumptions discussed in Sect. 2(c) are accurate. Thus, adopting the many-valued model may best approximate the world where everyone gets what she deserves.[47]

In conclusion, the retributivist aims to realise the apportioning of desert that is closest to the ideal situation where everyone is given what she deserves. It is plausible to argue that the many-valued model of the system of verdicts would be the best way to achieve this

---

[44] Thus, a probability indicating that the defendant is more likely guilty than not. Of course, if we inflicted a quantum of punishment when the probability of guilt is lower than .5, the probability that we would be punishing the non-deserving would be higher than the probability that we would be punishing the deserving.

[45] True, the punishments corresponding to the intermediate verdicts in the example of many-valued system given in the Introduction may not be ideal from a retributive perspective. This is because, as explained later in the article, they are optional and prospective. And yet, from a retributive perspective it would still be preferable to punish the guilty defendant with these measures than not to punish her at all.

[46] This may be true irrespective of whether the intrinsic evil consisting in acquitting a guilty person of a given crime has the same magnitude as the intrinsic evil consisting in punishing an innocent person for that very crime—an equivalence that probably all retributivists deny.

[47] Notice that the retributivist case for the many-valued model may be couched in decision-theoretic terms, in a way similar to what I have done with respect to consequentialism. A decision-theoretic analysis would probably lead to the conclusion that the many-valued model allows decision-makers to maximise the relevant conception of expected utility, according to which 'utility' consists in giving defendants what they deserve.

result. This argument does not seem to encounter any conceptual objection on the part of the retributivist,[48] only pragmatic challenges.[49]

3.3 Communicative Theories of Punishment

According to communicative theories, punishment is an act that expresses censure to the wrongdoer and at the same time tries to elicit from her a response, in the form of repentance, reform, and reconciliation.[50] In fact, not all communicative theories of punishment stress the importance of the wrongdoer's engagement. Some are eminently focused on the initial step of the communication, i.e., the expression of censure through punishment.[51] In what follows, I will concentrate on the most prominent and elaborated communicative theories advanced so far: Andrew von Hirsch's and Antony Duff's. Again, the question to address is whether a many-valued model of the system of verdicts may be compatible with these theories.

While both von Hirsch's and Duff's theories claim that punishment should perform a censuring function, they disagree on the appropriate role of hard treatment—i.e., the burden or pain that punishment normally involves. According to von Hirsch,[52] hard treatment plays primarily a preventive role. The criminal law aims at crime reduction. Censure may not be sufficient to dissuade people from committing crimes, because fallible

---

[48] Consider that a many-valued system need not conflict with the commitment not to punish identified innocent people intentionally or knowingly, which I mentioned earlier when discussing the fallibility argument. On the other hand, as seen earlier, the fact that innocent people would be convicted is not a prerogative of this system.

[49] Here I am not discussing the question as to whether the mixed theories of punishment elaborated by Hart (see Hart, *supra* note 8, in particular Chapter 1) and by Tadros (see V. Tadros, *The Ends of Harm* (Oxford University Press, 2011)) would be compatible with the many-valued model of the system of verdicts. It is my tentative conclusion that if these theories allow for a trade-off between the disvalue of not punishing those who should be punished and the disvalue of punishing those who should not be punished—as they must do if they aim at justifying a workable system of punishment—and if these theories accept a principle of proportionality—as they both do—then the many-valued model should be a viable option for them too, for the reasons explained above. Similarly, I suspect that the many-valued model of the system of verdicts may be compatible with Alan Brudner's 'legal retributivism' (see A. Brudner, *Punishment and Freedom: A Liberal Theory of Penal Justice* (Oxford University Press, 2009), Chapter 1). Under his theory, the purpose of punishment is to vindicate rights by showing the incoherency in the wrongdoer's initial denial of rights: she who denied rights has her rights denied through punishment; moreover, she can be taken to have authorised her own punishment through her initial act of denial. Acquitting the guilty is a failure to vindicate rights and thus an undesirable outcome. Punishing the innocent is a mere violation—rather than a vindication—of rights and thus an undesirable outcome. If the theory allows for a trade-off between the two—one seems inevitable if the theory aims at justifying a workable system of punishment—and if it demands proportionality—as it does—then the many-valued model may be a viable option under this theory too, for the reasons previously discussed.

[50] See R.A. Duff, *Punishment, Communication, and Community* (Oxford University Press, 2001), in particular, Chapter 3.

[51] See J. Feinberg, 'The Expressive Function of Punishment' in J. Feinberg, *Doing and Deserving: Essays in the Theory of Responsibility* (Princeton University Press, 1970) and A. von Hirsch, *Censure and Sanctions* (Clarendon Press, 1993). Fisher defends the compatibility of the many-valued model with expressive theories of punishment. However, her focus is on the message that the criminal law is expected to send to the community, rather than on the communicative interaction with the defendant. Also, Fisher does not elaborate on the formulation of intermediate verdicts—which is clearly an important issue for an expressive or communicative theory of punishment, as I will soon show. See Fisher, 'Conviction without Conviction,' *supra* note 13, at 862–871.

[52] This view is also defended in a work co-authored by Andrew Ashworth. See A. Ashworth and A. von Hirsch, *Proportionate Sentencing: Exploring the Principles* (Oxford University Press, 2005), Chapter 2.

agents—as we all are—may not be persuaded by moral reasons alone. Hard treatment supplies a prudential reason, which acts as a further disincentive to commit crimes.[53] However—von Hirsch is quick to stress—the censuring function 'has primacy' with respect to the purely preventive function: the latter can operate 'only *within* a censuring framework.'[54] According to Duff, instead, hard treatment performs a more complex role. Under his theory, punishment censures, but also aims at eliciting a qualified response from the wrongdoer, i.e., repentance, reform, and efforts to reconcile with the victims. This response may be achieved through censure alone. However, hard treatment is sometimes needed. Besides being a means of expressing censure, hard treatment 'provides a structure within which … [the wrongdoer] will be able to think about the nature and implications of his crime, face up to it more adequately than he may otherwise … do, and so arrive at a more authentic repentance.'[55] Reform—i.e., recognising the need to avoid criminal behaviour in the future, through internalising the relevant values—cannot be achieved without repentance, and thus may require hard treatment. Finally, also the reconciliation between the wrongdoer and the victim may be achieved through hard treatment, in particular if this has a reparative nature.[56] It should be pointed out that von Hirsch's and, especially, Duff's conceptions of the ideal forms of hard treatment depart significantly from some of the forms that hard treatment takes in current systems of punishment.

In the many-valued model, punishment is not exclusively a consequence of the guilty verdict. It also results from intermediate verdicts. In fact, if we were to follow a communicative theory of punishment, the intermediate verdicts themselves might be viewed as instances of punishment—as the guilty verdict is under such theories. These verdicts might be the primary means through which a message is sent to the defendant. But can intermediate verdicts communicate a message of censure, as von Hirsch and Duff would demand from punishment? Remember that these verdicts may only state a certain probability of guilt—whether this is expressed in numbers or words. They do not express a categorical statement to the effect that the defendant is or is not guilty. In fact, it is doubtful that we would be justified in expressing a categorical statement of guilt if the evidential threshold that was satisfied at trial were lower than proof beyond a reasonable doubt.[57] Also, if we were to brand and to treat the defendant as guilty when only intermediate thresholds are met, we would not really be implementing a many-valued system, but merely lowering the standard of proof of the current binary system. The problem is that we do not censure someone by telling her that there is a given probability that she has committed a crime. This would be taken as an assertion; probably as a warning or admonition; but not as an outright sign of disapproval. Nothing short of communicating the probability of 1—through a categorical statement of guilt—would do for the purposes of censure. Even though we can never be certain that the defendant committed the crime, if our aim is to censure her we must act as if we were. We must tell her that she is guilty—not just likely to be so. While intermediate verdicts are about the evidence and its assessment, guilty verdicts are about the event to which the evidence refers.[58] This is why the latter are apt vehicles for censure—not the former.

---

[53] See von Hirsch, *supra* note 51, at 12–13.

[54] *Id.* at 14.

[55] Duff, *supra* note 50, at 108.

[56] See *id.* at 109.

[57] See Picinali, *supra* note 11.

[58] Cf. C. Nesson 'The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts' (1985) 98 Harv. L. Rev. 1357.

Intermediate verdicts do not censure. However, if a guilty person were to receive one of these verdicts, it is possible that they may perform a function similar to censure. They may invite the guilty to focus on the nature and implications of her crime. This, coupled with the realisation of having just escaped censure and the most serious punishment, may potentially increase the feeling of guilt. Moreover, these verdicts may send powerful warnings to the effect that the guilty went very close to being found such, and that this may indeed happen on future occasions should she decide not to stay away from crime. Therefore, if repentance and reform are indeed among the purposes of punishment, intermediate verdicts may be well suited for these purposes. And, even if it were the case that intermediate verdicts cannot foster repentance and reform, it would still be plausible to argue that they could play an important role in fostering the internalisation of the values protected by the criminal law—which is *per se* a desirable result for communicative theories. As far as reconciliation is concerned, I admit that it is less likely that a victim would be willing to reconcile with the defendant if the latter is not publicly treated as guilty. And yet, through encouraging internalisation, repentance, and reform, an intermediate verdict may encourage the defendant to seek reconciliation privately. In any case, the fact that intermediate verdicts may not be successful in achieving all the three Rs[59] of punishment does not necessarily disqualify them. Whatever the extent of the positive result of intermediate verdicts, this result would hardly be achieved in a binary system of verdicts with a high standard of proof: in this system, the guilty defendants who are not proven guilty to the satisfaction of such a standard would all be acquitted.

Given that intermediate verdicts do not censure, it is not an argument against the many-valued model to claim that an innocent who were to receive an intermediate verdict would be wrongly censured. True, this innocent person would be subjected to hard treatment—e.g., in the forms previously described in my example of many-valued system. To the issue of hard treatment I now turn.

If we were to adopt von Hirsch's communicative theory, we would justify hard treatment by appealing to its preventive function. If so, we would find it difficult to justify the imposition of hard treatment in the case of intermediate verdicts. After all, von Hirsch tells us that the preventive function must have a secondary role, as it should only operate within a censuring framework. However, as I pointed out earlier, intermediate verdicts do not provide such a framework. This means that in the case of these verdicts punishment would perform primarily a preventive function—which is not to deny that it may also perform functions similar to censure, as argued above. It seems, therefore, that von Hirsch's theory—if interpreted strictly—is incompatible with the many-valued model of the system of verdicts, or at least incompatible with a many-valued system that attaches some form of hard treatment to the intermediate verdicts. To be sure—as Duff pointed out[60]—it is doubtful whether von Hirsch's own ideal system of punishment would be compatible with his requirement about the primacy of censure. This, however, is not the place to address this doubt.

Now suppose that we subscribed to Duff's theory. If hard treatment takes the forms previously described in our example of many-valued system, I doubt that we would have strong objections against inflicting it as the consequence of an intermediate verdict. To recapitulate, these forms are: the admissibility—with leave of the court—of the verdict as evidence of bad character in future proceedings for a similar offence; an exception to the double jeopardy guarantee (e.g., the test to be passed for a new trial for the same crime to

---

[59] This is how Duff refers to repentance, reform, and reconciliation. See Duff, *supra* note 50, at 107.

[60] See *id.* at 87–88.

be possible would be less stringent than the current 'new and compelling evidence' test); and an automatic sentence enhancement if the defendant is later convicted for another crime. Take first the case of the guilty defendant who receives an intermediate verdict. These forms of hard treatment are all 'prospective' and don't have an expiration date—although one may be added, if considered suitable. This means that they will accompany the guilty defendant for as long as she lives. Moreover, two of these treatments are in an important sense 'optional.' The defendant has a certain degree of control as to whether their detrimental consequences will be activated or not: if she decides to stay away from crime, it is unlikely that they will. The duration of these treatments, coupled with the appeal that they make to the defendant's responsibility, are likely to reinforce the message sent by the intermediate verdict, and thus to achieve the desiderata of punishment. Importantly, this is a result that a binary system of verdicts with a high standard of proof would hardly obtain, given that in such a system the guilty defendants at issue here would all be acquitted. What about the innocent defendants who receive an intermediate verdict with the corresponding hard treatment? True, in a binary system with a high standard of proof they would have simply been acquitted. But would they face a hard time under a many-valued system as the one in my sketch? My tentative contention is that probably they wouldn't. These innocent people would not be the targets of censure: as seen earlier, intermediate verdicts reflect the evidence but do not censure the conduct. Also, given the prospective and optional nature of the specified hard treatments, innocent defendants would not have strong reasons for concern. Finally, the scenario consisting in the innocent defendant receiving an intermediate verdict would be sufficiently rare if the corresponding probability threshold is sufficiently high—and, in any case, higher than .5.

In conclusion, the supporter of a communicative theory of punishment may agree with this—by now familiar—statement: the additional harm that the many-valued model produces by punishing innocent people through intermediate verdicts would be outweighed by the model's ability to reduce the harm consisting in the acquittal of the guilty—a statement whose accuracy depends on whether the (reasonable) assumptions discussed in Sect. 2(c) are themselves accurate. If the arguments presented in this section are sound, it seems that not even the supporters of communicative theories of punishment need to object as a matter of principle to a many-valued model of the system of verdicts.

4 Conclusion

In this essay, I have tried to show that consequentialist, retributive, and communicative theories of punishment are not in principle incompatible with—and in fact may be well served by—the many-valued model of the system of verdicts. Whether a specific many-valued system is compatible with these theories will depend on how it is designed: in particular, on how high are the evidential thresholds corresponding to the intermediate verdicts and on the features of the respective punishments. However, the many-valued nature of the system doesn't seem to be itself a reason for concern. As I have attempted to show, a many-valued system such as that sketched in the Introduction seems immune from in-principle objections, and may outperform the current binary system irrespective of which of the above theories of punishment one were to adopt. Or, at least, it may do so if the following (reasonable) assumptions are true: (1) that the majority of defendants who go to trial are guilty; and (2) that the evidence that is presented in trials of guilty defendants is generally stronger (and thus indicates a higher probability of guilt) than the evidence that is

presented in trials of innocent defendants. Were these assumptions false, we may doubt that such many-valued system would yield mistakes that in the aggregate are less significant than those produced by a binary system—where the significance of a mistake is measured according to the criteria that are relevant for each theory of punishment. However—as argued earlier—if either of these assumptions were false, our criminal justice system would hardly be justified, and would need much more incisive reform than a switch to a many-valued model.

The fact that the many-valued model may be compatible with consequentialist, retributive, and communicative theories of punishment does not rule out that it may be better suited to achieve the goals of one of these theories rather than those of the other two. Also—given that these are really families of theories rather than individual theories—this fact does not rule out that the many-valued model may suit one theory within the family better than others—as we have already seen in the discussion of communicative theories of punishment. These suitability issues will be best addressed once the pragmatic part of the project is underway. In other words, when we will have more information on how the many-valued model performs, or can be expected to perform.