

CRIMINAL RESPONSIBILITY FOR UNSPECIFIED OFFENSES

*Alon Harel & Ariel Porat**

Should a court convict a defendant for unspecified offenses if there is no reasonable doubt that he committed an offence, even though no particular offence has been proven beyond reasonable doubt? Suppose a defendant is charged with two unrelated offenses, for example, pick-pocketing and rape, allegedly committed at different times and places. The probability that he committed each one of the offenses is .9. Assume that the minimum threshold required for conviction is .95. Under prevailing evidence law, the defendant would be acquitted of both charges since no offense can be specifically attributed to him. However, a simple calculation of the probability that the defendant committed at least one offense amounts to .99. Consequently, it seems that convicting him for one offense without specifying what this offense is and punishing him with the sanction designed for the least severe offense would be just and efficient. We call the principle that requires such an aggregation the "Aggregate Probabilities Principle" (APP).

This Article establishes that under certain conditions, deterrence, efficient law enforcement, and minimization of adjudication errors would be better achieved were courts to apply such an APP. The Article also addresses the most powerful possible objections to this method and

* Alon Harel is Phillip P. Mizock & Estelle Mizock Professor in Administrative and Criminal Law, at the Hebrew University of Jerusalem and a Visiting Professor of Law at University of Texas Law School (Spring 2008). Ariel Porat is the Alain Poher Professor of Law at Tel Aviv University and Fischel-Neil Distinguished Visiting Professor of Law at the University of Chicago. For helpful comments and discussions, we thank Abraham Bell, Meir Dan-Cohen, Massimo D'Antoni, Rosalind Dixon, Shai Dotan, Frank Easterbrook, David Enoch, Talia Fisher, Oren Gazal-Ayal, David Gilo, Gabriel Hallevy, Bernard Harcourt, Todd Henderson, Assaf Jacob, Leo Katz, Shai Lavi, Shmuel Leshem, Saul Levmore, Orly Lobel, Andrei Marmor, Richard McAdams, Jacob Nussim, Timna Porat, Eric Posner, Amit Pundik, Joseph Raz, Eli Salzberger, Chris Sanchirico, Andrew Simester, Alex Stein, Lior Strahilevitz, and Avraham Tabbach. We thank also the participants in the 7th Annual Meeting of the Israel Law and Economic Association, the 2008 Annual Meeting of the European Law and Economic Association held in Haifa, as well as the participants in workshops at Bar-Ilan University, Haifa University, the Interdisciplinary Center in Herzliyah, Ono Academic College, the Rationality Center at the Hebrew University and Tel Aviv University, and in the Siena-Tel-Aviv-Toronto Law and Economics Workshop held in Tel Aviv in 2008 and the law and economics workshops at the University of Chicago and the Hebrew University. Our thanks to Roni Schocken for his superb research assistance and Dana Rothman-Meshulam for excellent language editing.

suggests that the aggregation principle be adopted only under strict conditions that preclude its potential abuse by the prosecution. In addition, the Article shows that sometimes aggregating probabilities will yield less, rather than more, convictions. If the APP is adopted, the presumption of innocence currently applied with regard to the offense will be replaced by a presumption of innocence applied to the accused.

TABLE OF CONTENTS

INTRODUCTION	2
I. INTRODUCING THE AGGREGATE PROBABILITIES PRINCIPLE	6
II. THE AGGREGATE PROBABILITIES PRINCIPLE IN CONTEXT	10
<i>A. Aggregating Probabilities under Prevailing Law.....</i>	<i>10</i>
<i>B. Pattern of Behavior and Interdependence.....</i>	<i>16</i>
1. Prior Acts and Similar Crimes Doctrines	16
2. Interdependence.....	18
III. THE CASE FOR THE AGGREGATE PROBABILITIES PRINCIPLE.....	22
<i>A. Adjudication Errors.....</i>	<i>22</i>
<i>B. Deterrence.....</i>	<i>24</i>
<i>C. Costs of Enforcement.....</i>	<i>26</i>
IV. PRACTICAL OBJECTIONS	27
<i>A. Manipulations by the Prosecution and Agency Costs.....</i>	<i>27</i>
<i>B. Statistical Evidence and Statistical Inferences.....</i>	<i>29</i>
<i>C. Increased Litigation.....</i>	<i>31</i>
<i>D. Implementation Difficulties and the Meaning of a Probabilistic Threshold.....</i>	<i>32</i>
<i>E. Redundancy.....</i>	<i>34</i>
V. RETRIBUTIVIST AND EXPRESSIVIST THEORIES OF PUNISHMENT...	36
<i>A. The Case against Aggregation of Probabilities.....</i>	<i>36</i>
<i>B. The Case for a Moderate Aggregate Probabilities Principle</i>	<i>40</i>
CONCLUSION	41

INTRODUCTION

The question we address in this Article is the method by which courts¹ should evaluate whether the standard of “beyond reasonable doubt” is satisfied in cases in which a person is (or has been) charged with several offenses. Under the traditional approach, when a defendant is charged with a number of offenses, the court should examine each charge individually

¹ The phrase “court” is used in this paper for both judge and jury.

and decide whether the “beyond reasonable doubt” standard is satisfied *with regard to each charge*. In contrast, we suggest here that, in addition to considering each charge separately, the court should examine all charges in aggregate and decide whether the standard is satisfied with respect to *at least one* of them. We call the former approach the “Distinct Probabilities Principle” (DPP) and the latter the “Aggregate Probabilities Principle” (APP). Example 1 illustrates how the APP would work.

Example 1. Unrelated Offenses (1). A person is charged with two unrelated offenses allegedly committed by him at different times and places: pick-pocketing and rape. The evidence suggests that the probability that he committed each one of these offenses is .9. Assume that the required probability necessary to satisfy the beyond reasonable doubt standard is .95.² Should the court convict the defendant on any of the offenses?

Weighed separately—that is, applying the DPP—the defendant ought to be acquitted of both offenses. Yet it is also easy to see that the probability that he committed at least one of these offenses is higher than the probability necessary for conviction in a criminal trial. In fact, the former probability is .99.³ Consequently, it could be argued that it is not unjust to use the APP and convict him of one unspecified offense and impose on him *at least* the sanction set for the least severe of the two offenses, i.e., pick-pocketing. Note that if the defendant had been charged with four instead of two offenses, the probabilistic calculation would yield a probability of .9999 that he had committed at least one offense, and applying the APP would guarantee conviction. In fact, in such a case, a conviction for two offenses would be warranted since the probability that two offences had been committed would be higher than the threshold required for conviction.⁴

² This assumption does not require that we presuppose the beyond reasonable doubt standard to be exclusively probabilistic, i.e., as resting exclusively on probability. We can assume that a certain amount of individualized (rather than statistical) evidence relating to the defendant's conduct is a prerequisite for conviction. For using the .95 threshold as an illustration of the probabilistic nature of the beyond reasonable doubt standard, see David Kaye, *Laws of Probability and the Law of the Land*, 47 U. CHI. L. REV. 34, 40 (1979) (“Surely it is not some defect in probability theory that restrains us from instructing jurors that they should convict so long as they are, say, at least ninety-five percent certain that the defendant is guilty.”)

³ Here is the calculation: the probability that the defendant committed each one of the offenses is .9, and therefore the probability, for each one, that he did not commit the offense is $1 - .9 = .1$. Consequently, the probability that he did not commit any offense is $(.1)^2 = .01$, and the probability that he committed at least one of the offenses is $1 - .01 = .99$.

⁴ This is the outcome of a binomial distribution. There are 4 events, and in each one, the defendant either committed the offense or did not (thus he either committed 0, 1, 2, 3, or 4 offenses, and the probability that one of these scenarios transpired is 1). To calculate the probability that the defendant committed at least 2 of the 4 offenses, the probability that he committed 0 offenses or 1 offense should be subtracted from 1. Since the probability that the defendant did not commit any offense is $(.1)^4 = .0001$, and the probability that he committed

The dilemma that this Example raises is straightforward. If the legal system applies only the DPP and, consequently, the defendant in Example 1 is acquitted of all offences, he will escape conviction despite the fact that the probability that he committed at least one offense is greater than the probability required for criminal conviction. Individuals are routinely convicted for committing a single offense on the basis of evidence that establishes guilt with a lower probability than the aggregate probability that arose from the evidence in example 1. Under a probabilistic calculation and applying the APP to our case would lead, however, to the outcome that the defendant cannot be convicted of any *specific* offense. All that would be established by the evidence is that he committed “beyond reasonable doubt” at least one offense out of the two with which he was charged. Is it just or efficient that a defendant in such a case is acquitted while, at the same time, we convict a person who is charged with a single offense that can be proven at a lower probability?

Example 1 illustrates how aggregating the probabilities of all charges (in accordance with the APP) results in more convictions than when each charge is examined separately (in accordance with the DPP). The next example illustrates how aggregating probabilities could result in fewer convictions.

Example 2. Unrelated Offenses (2). A person is charged with two unrelated offenses allegedly committed by him in different times and places: pick-pocketing and rape. The evidence suggests that the probability that he committed any one of these offenses is .95. Assume that the required probability necessary to satisfy the beyond reasonable doubt standard is .95. Should the court convict the defendant on both offenses?

If the offenses are examined independently, the defendant ought to be convicted on both charges. Yet it is easy to see that the probability that the defendant committed *both* offenses is lower than .95. In fact, that probability amounts to about .9!⁵ The court in this case would therefore be faced with a difficult choice. It could convict the defendant of one offense, since the probability that he committed at least one offense is greater than .95 (it is .9975) and thus satisfies evidence law requirements. Yet, if the court did so, it could not specify the offenses for which the person is

exactly 1 offense is $(.9) * (.1)^3 * 4 = .0036$ (.9 is the probability that he committed one specific offense; $(.1)^3$ is the probability that he did not commit any of the other 3 offenses; we multiply by 4 because the specific offense committed by the defendant could be any of the 4 offenses), the probability that he committed at least 2 offenses is $1 - .0001 - .0036 = .9963$. To calculate the probability that the defendant committed at least 3 offenses, the probability that he committed 4 offenses should be added to the probability that he committed 3 offenses. Since the probability that the defendant committed 4 offenses is $(.9)^4$ and the probability that he committed 3 offenses is $(.9)^3 * .1 * 4$, the probability that he committed at least 3 offenses comes to $.6561 + .2916 = .9477$.

⁵ $(.95)^2 = .9025$.

convicted. Instead, the court would simply base the conviction on the ground that the probability that the defendant committed at least one of the two offenses is high enough to justify his conviction. In contrast, if the court were to decide to convict the person of both offenses, it would result in a conviction despite the fact that there is less than .95 likelihood that he committed both offenses. Even though similar rationales support the application of the APP to cases represented by Examples 1 and 2, the focus of this paper is on cases where the Principle would result in more, rather than fewer, convictions—namely, those cases represented by Example 1.

Surprisingly, the possibility of using the APP in criminal law has yet to be explored.⁶ To the best of our knowledge, the Principle has never even been discussed or considered in case law and we know of no case in which a prosecutor or a defense lawyer has suggested applying it. It seems that lawyers as well as theorists take it for granted that a person can only be convicted for committing a specific identifiable crime.

We find this perplexing because, as we will show below, both justice and efficiency considerations support applying the APP to a broad range of cases. This Article aims at exploring this puzzle. It explores and addresses the most powerful objections to applying the APP in prevailing criminal law and proposes that certain conditions be set for the adoption of the principle that would preclude its abuse by the prosecution. If the APP is adopted as advocated, the presumption of innocence currently applied with regard to the offense will be replaced by a presumption of innocence applied to the accused: if there is no reasonable doubt that the accused committed some unidentifiable offence, she ought to be convicted of committing an offence even if no particular offence has been proven beyond reasonable doubt.

The Article proceeds as follows. *Part I* introduces the Principle and explores its potential scope. It also contrasts the applicability of the APP in cases in which the different offenses attributed to the defendant are of completely different natures (as in Example 1) with its applicability to cases of identical offenses. *Part II* distinguishes the APP from other types

⁶ In other fields, however, the APP has been considered and discussed at length; for instance, legal writers have proposed aggregating probabilities in *civil* cases, *see, e.g.*, Saul Levmore, *Conjunctions and Aggregation*, 99 MICH. L. REV. 723 (2000-2001) (discussing aggregating probabilities mainly in tort cases). Furthermore, in an insightful paper published over a decade ago, Frederick Schauer and Richard Zeckhauser proposed aggregating probabilities across cases outside the judicial context, Frederick Schauer & Richard Zeckhauser, *Degree of Confidence for Adverse Decisions*, 25 J. LEGAL STUDIES 27, 41-51 (1996). The authors argue that it would make sense for a school to dismiss a teacher against whom several complaints of sexual harassment had been made in the past, even if each complaint, considered separately, would not constitute sufficient reason for dismissal. The authors assume, however, that such an argument is inapplicable to criminal proceedings. “Of course, the practice of noncumulation of charges in the criminal law serves important goals ... Obviously there are costs associated with these goals ... but weighing the costs and benefits of the refusal to cumulate in the criminal process is not our goal.” *Id.* at 45-46.

of aggregations of probabilities conducted in both criminal and tort law. It shows that the only doctrine that comes close to the APP is Market Share Liability, which is recognized in torts by several jurisdictions. In this Part, we also explore under what conditions interdependence between the offenses attributed to the defendant should preclude the application of the APP. We show that in most of the cases, regardless of the interdependence of the offenses, the APP should apply. In addition, a comparison to the Prior Acts and Similar Crimes doctrines is presented in this Part in order to distinguish the APP from other, related doctrines grounded on completely different rationales. In *Part III*, we argue that deterrence, efficient law enforcement, and minimization of adjudication errors would be better achieved under the APP than under the DPP. We also show that the APP would not inevitably increase the number of wrongful convictions; if applied in conjunction with a higher conviction threshold than currently prevailing, the number of total convictions will remain unchanged but there will be a consequent reduction in the number of false convictions. *Part IV* raises several possible practical objections to the APP, the most important being its potential abuse by enforcement agencies (the police and prosecution) and difficulties of implementation. This Part shows how none of these objections is compelling enough to justify rejecting the APP outright, although some do mandate particular caution in applying the Principle.

Finally, *Part V* analyzes the APP from a justice-based perspective. It establishes that whereas retributivists (or, more accurately, proponents of certain versions of retributivism) might be inclined to impose sanctions on a defendant for *an* offense committed irrespective of whether the offense can be specified (and, consequently, would adopt the APP), expressivists might contend that sanctions are justified only if it has been shown that the defendant committed a *specific, well-defined* offense (and, consequently, would adopt the DPP). It is therefore expressivism that explains why criminal law is currently unwilling to use the APP. Yet, we maintain that under certain conditions the defendant may be convicted on the basis of the APP without undermining the expressivist concerns of punishment. Conclusions follow.⁷

I. INTRODUCING THE AGGREGATE PROBABILITIES PRINCIPLE

To convict a person in a criminal trial, the prosecution must prove the charges “beyond reasonable doubt.” The rationale underlying this

⁷ The APP could be applied also across civil cases. The considerations for and against such application differ from those relevant to criminal cases, and we leave it to future consideration.

requirement and its precise meaning, are, of course controversial.⁸ Yet, it is undisputed that the standard has an important probabilistic aspect to it:⁹ the evidence that grounds a conviction in a criminal trial must establish that the defendant committed an offense at a high degree of probability.¹⁰

It is implicitly assumed in criminal and evidence law that in order to convict a person for committing an offense, the beyond reasonable doubt standard should be met *with respect to each particular offense*, that is, separately and distinctively from other offenses.¹¹ To the best of our

⁸ See, e.g., Lawrence M. Solan, *Refocusing the Burden of Proof in Criminal Cases: Some Doubt about Reasonable Doubt*, 78 TEXAS L. REV. 105, 105 (2000) (“Most debate in judicial opinions and in the scholarly literature has focused on whether reasonable doubt should be defined for the jury, and, if so, how it should be defined.”); Note, *Reasonable Doubt: An Argument Against Definition*, 108 HARV. L. REV. 1955, 1955 (1995) (concluding that “courts should not attempt to define the term [reasonable doubt] in conveying the reasonable doubt concept to juries”); Jessica N. Cohen, *The Reasonable Doubt Jury Instruction: Giving Meaning to a Critical Concept*, 22 AM. J. CRIM. L. 677, 678 (1995) (arguing that “because reasonable doubt is a term of art it should be defined for the jury”); Henry A. Diamond, Note, *Reasonable Doubt: To Define, or Not to Define*, 90 COLUM. L. REV. 1716, 1716 (1990) (“jury instructions defining reasonable doubt should always be given in criminal trials and are constitutionally required when requested by the defendant or the jury”). See also Thomas V. Mulrine, *Reasonable Doubt: How in the World Is It Defined?*, 12 AM. U. J. INT’L L. & POL’Y 195, (1997) (explaining various approaches to and definitions of reasonable doubt).

⁹ Alex Stein, FOUNDATIONS OF EVIDENCE LAW 65 (2005) (“Adjudicative fact-finding rests on probabilistic reasoning that derives from experience.”); *id.* at 66 (“Any finding that fact-finders make can only be probable, rather than certain.”); Jonathan J. Koehler & Daniel N. Shaviro, *Verdical Verdicts: Increasing Verdict Accuracy Through the Use of Overly Probabilistic Evidence and Methods*, 75 CORNELL L. REV. 247, 252 (1990) (“All evidence is probabilistic, in the sense that there is a risk of error in relying on it to support a factual conclusion about a case.”).

¹⁰ For those readers who are skeptical about mathematical calculations in the legal context, we suggest considering the same problem without resorting to probabilities: Should a court convict a defendant when there is no reasonable doubt that he committed at least one of several charged offenses, but it cannot be established which was specifically committed by him? See L. Jonathan Cohen, *The Role of Evidential Weight in Criminal Proof*, 66 B.U. L. REV. 635, 635 (1986) (stating that by trying to give an account of the standard of criminal proof in Pascalian terms, one reserves the crucial place in reasoning for the assignment of a high value *non*-Pascalian function for the assessment of evidential weight); Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1372-75 (1971) (hereinafter: Tribe, *Trial by Mathematics*) (stating that “[b]oth callousness and insecurity ... might be increased by the explicit quantification of jury doubts in criminal trials—whether or not it would be *factually accurate* to describe the trial system as imposing criminal sanctions in the face of quantitatively measured uncertainty in particular cases”); Laurence H. Tribe, *A Further Critique of Mathematical Proof*, 84 HARV. L. REV. 1810, 1815-17 (stating that the use of Bayesian methods in criminal trials would undermine the presumption of innocence); Charles R. Nesson, *Reasonable Doubt and Permissive Inferences: The Value of Complexity*, 92 HARV. L. REV. 1187, 1225 (1979) (concluding that “any conceptualization of reasonable doubt in probabilistic form is inconsistent with the functional role the concept is designed to play”).

¹¹ The Prior Acts doctrine and the Similar Crimes doctrine can be regarded as deviating from this principle. However, under closer scrutiny, neither of these two doctrines appears to be a true deviation from the general principle. Both are invoked to establish that the

knowledge, this principle, the Distinct Probabilistic Principle (DPP), has never been questioned. The Aggregate Probabilities Principle (APP), which we propose adopting, challenges the DPP: Why should we not convict a person for *an* offense when it is certain, or almost certain, that that person committed some offense, even if it cannot be established which one? Why should we not convict the defendant in Example 1 for at least one offense when the probability that he committed no offense at all is 1%? Or, if we modify example 1 to encompass four, instead of two, offenses, should the defendant be acquitted of all charges when the probability he committed no offense at all is *one in ten thousand*? To reduce the risk of excessive punishment, proponents of the APP would maintain that the defendant in Example 1 should be punished with the degree of severity that the least serious of the two offenses would warrant.

Note, as we asserted earlier, that the APP works not only against defendants but also in their favor. Example 2 is illustrative of this: under the APP, a defendant accused of two offenses where each offense—if examined separately—can be proven beyond reasonable doubt should not be convicted for both but, rather, only for one.¹²

In theory, the APP could be applied to any case in which a defendant is charged with more than one offense. Below, we draw some intuitive distinctions between different types of cases, returning to these distinctions in the ensuing Parts to show their potential normative significance in shaping the APP.

First, it is intuitively sound to distinguish cases in which a person is charged with identical offenses from cases in which a person is charged with different offenses ("*the nature of the offense criterion*"). As we demonstrate later on, expressivist theories of punishment would find it more tolerable to convict a person for sexual assault when two charges of that same nature have been brought against him, even if none of them, if examined separately, can be proven beyond reasonable doubt. Expressivists would be reluctant to convict a person charged with pick-pocketing or rape when none of these offenses can be proven beyond reasonable doubt, even if it is evident that he committed one of them.

probability that the defendant committed a particular offense is higher than otherwise thought. In other words, these doctrines are used to help the decision-maker determine the probability that the defendant committed a *particular* offense. They are, therefore, not exceptions to the general principle prevailing in criminal law under which the legal system investigates exclusively the probability of *particular identifiable offenses* having been committed.

¹² This is not to say, however, that the APP is neutral overall with respect to defendants, i.e., that there is an identical number of expected convictions and expected acquittals resulting from applying the APP. The transition from the DPP to the APP can be expected to bring about more convictions than acquittals, based on the intuition that the APP, in taking into account all probabilities from 1% to 94% (assuming 95% is the threshold for conviction), increases the number of convictions, and only in taking into account probabilities from 95% to 99% does it reduce the number of convictions.

Second, even if the offenses are identical, it would be easier to accept conviction for what we label “*homogenous* offenses,” namely, offenses whose nature and severity are less dependent on the particular circumstances, than to accept conviction for “*heterogeneous* offenses” (“*the homogeneity criterion*”). The severity of rape or murder depends on numerous contextual considerations, whereas pick-pocketing or breaching the statutory speed limit is typically less a product of circumstances.

Third, in some cases, the relevant offenses were directed at the same victim, whereas in other cases, different victims were the targets of the different offenses (“*the same victim criterion*”). Thus, there could be a difference between applying the APP to a case in which an employee is accused of two thefts directed at his employer and applying it to the case of a defendant accused of two such acts targeting different victims. This differentiation could cut both ways: On the one hand, it could be more acceptable to convict a person for an unspecified offense if it can be proven beyond reasonable doubt that an offense was committed against a single victim than if an offense was committed against different victims. Some of the expressivist theories may provide support for such an argument. On the other hand, when there is a single victim of all the alleged offenses, the risk of interdependence of the charges that could preclude the use of the APP is greater. Thus, there could always be a concern that the employer, the alleged victim, actually sought to frame the accused leading to reasonable doubt with respect to the latter's guilt in each one of the charges. As we elucidate further on, interdependence could be sometimes a significant obstacle for the use of the APP.

Fourth, when we move from the core offenses of criminal law to regulatory violations, the use of the APP seems more reasonable. Regulatory offenses are governed primarily by considerations of deterrence, and justice-based considerations are less applicable with respect to these offenses (“*the regulatory offense criterion*”). For example, being caught speeding twice by a police radar seems to present a stronger case for applying the APP than committing two thefts. Moreover, in addition to being regulatory violations, traffic offenses typically satisfy the homogeneity criterion.

Fifth, there is an intuitive difference between cases in which the defendant is charged with all offenses simultaneously and those in which he is charged with a new offense after having been previously convicted on or acquitted for other offenses (“*the same trial criterion*”). Compare Example 1, where there are two simultaneously charged offenses and the evidence suggests that the probability that the defendant committed each of the offenses is .9, with a case in which a person has been acquitted once in the past because the probability that he committed the past offense was only .9. Similarly, consider Example 2, in which the defendant is charged with two offenses, each of which can be proven with a probability of .95, compared to a case in which the defendant was convicted once in the past because the evidence indicated a probability of .95 that he had committed

the given offense. Somewhat counter-intuitively, applying the APP would yield higher chances of conviction for a person acquitted in the past and lower chances of conviction for someone convicted in the past. Yet the case for applying the APP across different trials, rather than different charges within the same trial, is weaker, and we thus do not advocate it. The obstacle to the use of the APP across different trials derives in part from implementation considerations and, in part, from principles of the finality of judicial decisions.¹³

Finally, the APP is not limited to cases in which the product of the aggregated probabilities is less than 1. Rather, it could also apply to cases in which there is no doubt whatsoever that the defendant committed an offense, even though it cannot be established which offense. Leo Katz has offered an illustrative example of such a case: Suppose a murder and a burglary were committed at the same time in two different places, and hidden cameras recorded both incidents. Unfortunately (for the law enforcement authorities), the perpetrators of these crimes are twin brothers. It is known, therefore, that each of the two brothers committed one of the offenses. It is unknown, however, which offense was committed by which brother.¹⁴ Under the APP, both brothers would be convicted for the lesser of the two crimes, namely, burglary.

II. THE AGGREGATE PROBABILITIES PRINCIPLE IN CONTEXT

A. *Aggregating Probabilities under Prevailing Law*

Aggregating probabilities is not an unfamiliar phenomenon in the legal system. The issue of aggregating probabilities takes place in fact-finding procedures when a court or jury must determine whether a conjunction of facts or events transpired. For each of the facts or events comprising the set, there is a specific probability that it took place, and the probability that *all* the facts or events or, alternatively, at least one of them took place is an aggregation of all relevant probabilities. Thus, suppose a judge in a civil case must decide whether the defendant was negligent and whether he

¹³ *First*, the information obstacles in applying the APP across trials are more serious than those that would arise across charges in the same trial. *See infra* Section IV.D. *Second*, taking into account prior acquittals as a consideration for convicting the same defendant in a subsequent trial could violate the Fifth Amendment “double jeopardy” clause. *Cf. infra* note 30 and accompanying text. *Third*, applying the APP across different trials in cases represented by Example 2 (when the APP generates fewer, rather than more, convictions) could inefficiently reduce deterrence of future crimes: a defendant who was convicted in a trial for one offense at a probability of .95 will not be punished for a subsequent crime as long as the prosecution cannot establish his guilt at a probability of 1. *See infra* Section III.B.

¹⁴ LEO KATZ, *ILL-GOTTEN GAINS: EVASION, BLACKMAIL, FRAUD, AND KINDRED PUZZLES OF THE LAW* 67-69 (1996).

caused the given injury, and only if both questions are answered affirmatively the defendant is found liable. Assume now that the judge estimates the probability of the defendant's negligence at .6 and the probability that, given this negligence, he caused the injury at .6 as well. Aggregating the probabilities in this case yields a probability of .36 that the defendant was both negligent *and* caused the injury (what we will refer to as "*the civil cumulative case*"). If the judge's decision is based on an aggregation of the different probabilities of the two components of the wrongdoing, the plaintiff will lose since the probability of both negligence and causation amounts to only .36, which does not satisfy the "preponderance of the evidence" requirement. However, if each component of the cause of action is considered separately, the plaintiff will win, since the probability of each component amounts to .6. Legal theorists disagree as to whether an aggregation of probabilities rule should be applied in such a case, and it appears that case law is not inclined to use aggregation in such cases.¹⁵

Similar difficulties arise in cases of disjunctive, as opposed to conjunctive, liability. In the former, the defendant is liable if either scenario A took place *or* scenario B took place. Assume, for instance, that scenario A involves one form of negligence that could have caused the given injury and scenario B involves a different form of negligence that could have caused that same injury. Assume that the probability of scenario A is .3 and the probability of scenario B is also .3 and that these probabilities are independent of one another. The defendant is liable under either scenario A or scenario B ("*the civil alternative case*"). In such a case, the probability

¹⁵ See Levmore, *supra* note 5, at 752 n.58 (arguing that no jurisdiction explicitly recognizes the "product rule"—which is the rule that mandates the aggregation of the probabilities—and explaining that such non-recognition could be warranted mainly in those cases where decisions are made by either a jury or another multimember panel, either unanimously or by supermajority). Some theorists argue that reluctance to use the product rule generates "the conjunction paradox." See, e.g., Alex Stein, *Of Two Wrongs That Make a Right: Two Paradoxes of the Evidence Law and Their Combined Economic Justification*, 79 TEXAS L. REV. 1203-05 (2001) (identifying the distortion that is created by disallowing the product rule, but arguing that taken together with another major distortion in fact-finding, a second best solution is achieved). Allowing the product rule is supported by Maya Bar-Hillel, *Probabilistic Analysis in Legal Factfinding*, 56 ACTA PSYCHOLOGICA 267 (1984) (saying that "the conjunction of a small number of weakly probative characteristics can be strongly probative"); Kaye, *supra* note 2, at 55; Bernard Robertson & G.A. Vignaux, *Probability—The Logic of the Law*, 13 OXFORD J. LEGAL STUDIES 457 (1993) (concluding that "[o]nce one regards probability as a generalization of logic and has freed one's mind from the shackles of frequentist examples and the Mind Projection Fallacy, these objections [to the use of probabilities] evaporate. The logical rules for thinking about facts in legal cases are those of probability"). Disallowing the product rule is supported by Ferdinand Schoeman, *Cohen on Inductive Probability and the Law of Evidence*, 54 PHI. SCI. 76, 80-82 (1987) (discussing the conjunction paradox and claiming that mathematical probability is inadequate as a model for rational thinking). See also Barbara White, *Coase and the Courts: Economics for the Common Man*, 72 IOWA L. REV. 577 (1987) (claiming that the product rule masks a value system rather than providing an unbiased approach to dispute resolution).

that at least one of the scenarios occurred amounts to .51.¹⁶ Once again aggregating the probabilities would generate a different decision than when each probability is considered separately.¹⁷ While courts generally tend to reject the aggregation principle in civil alternative cases,¹⁸ there are some important exceptions to this. In tort, for instance, proving by a preponderance of the evidence that the defendant's wrongful act caused the injury in question would normally be sufficient for courts to impose liability, even if the plaintiff cannot prove by the same standard what *precisely* made the defendant's act wrongful.¹⁹

Finally, aggregating probabilities can also be relevant in the context of criminal cases. When establishing the defendant's guilt requires the proving of several components of the offense, aggregating the probabilities of each component separately will generate a different outcome than an integrated aggregation of the probabilities. Thus, for instance, it is possible that even if each component of the offense can be proven beyond reasonable doubt, reasonable doubt exists with respect to the cumulative presence of all components ("*the criminal cumulative case*"). Will the court convict the defendant under such circumstances? The answer to this question is unclear.²⁰

The case discussed in this article differs substantively from both the civil and criminal cumulative cases. The two latter cases relate to the

¹⁶ The probability that none of the events took place is $.7 \times .7 = .49$. The probability that *at least* one of them took place is $1 - .49 = .51$. If the scenarios exclude one another, then the probability that at least one took place is $.3 + .3 = .6$.

¹⁷ See Levmore, *supra* note 5, at 729 n.11, 745-46 (explaining the "alternative Routes" scenario and labeling this issue "reverse conjunction"). Levmore uses his explanation for the rejection of the product rule also for the civil alternative case.

¹⁸ Levmore, *supra* note 5, at 729 n.11 (stating that courts do not apply the product rule); Tribe, *Trial by Mathematics*, *supra* note 10, at 1361 (claiming that hard statistical data lead decision-makers to "dwarf the soft variables" and to assume that "[i]f you can't count it, it doesn't exist"). For a different view, see Koehler & Shavero, *supra* note 9, at 265 (stating that psychological research does not support Tribe's assumption, but, rather, "suggests that, in a wide range of situations, people generally undervalue base rate evidence and attach too much weight to case-specific evidence").

¹⁹ DAN DOBBS, *THE LAW OF TORTS* § 154, at 370-73 (2000) (describing cases where the jury is permitted to infer that the defendant was negligent in causing the harm in a specific scenario, even though there are several possibilities of a negligent act and none of them is sufficient on its own to warrant such an inference).

²⁰ See Levmore, *supra* note 5, at 729 (presenting a possible application of the product rule to criminal cases); *id.* at 733 n.19 (suggesting that the defense might benefit from a rule of aggregation when it reminds the jury of all the doubts that have been raised and implies that, combined, they create more than a reasonable doubt). See also Jonathan Remy Nash, *A Context-Sensitive Voting Protocol Paradigm for Multimember Courts*, 56 STAN. L. REV. 75, 138 (2003) (discussing the rule of aggregation in the context of voting by judges in a panel or by jurors and observing that "[a]lthough a criminal defendant cannot be convicted unless a jury unanimously finds each element of the crime charged proven beyond a reasonable doubt, 'a federal jury need not always decide unanimously which of several possible sets of underlying brute facts make up a particular element, say, which of several possible means the defendant used to commit an element of the crime.'" (citations omitted)).

question of when courts ought to convict a defendant for a particular specific offense or find him liable for a particular specific wrong. The aggregation of probabilities in such cases is aimed at determining whether the person committed the particular offense or the wrong attributed to them. A conviction or a finding that the defendant is liable implies that the court was satisfied that the evidence was sufficient to justify conviction or attribution of liability for a *particular* act. Conversely, an acquittal or a finding that the defendant is not liable implies that the court had not found the evidence to be sufficient to justify conviction or imposition of liability for a *particular* act. In contrast, this article seeks to examine cases in which no specific offense can be attributed to the defendant—those cases in which it is evident (or at least sufficiently probable) that the defendant committed an offense, but it is unclear what that offense is. Indeed, as opposed to the civil and criminal cumulative cases, the APP advocated in this Article is not about fact-finding but, rather, about *substantive* criminal law. We do not argue that courts are wrong in establishing whether an event occurred or not; instead, we maintain that the event courts should look at, and which they currently ignore, is the *commission of an unspecified offense*. Our argument touches on the basis of criminal liability, not the particularities of fact-finding.

The civil alternative case discussed above is more relevant to our inquiry. In both the civil alternative case and in our case, aggregating the probabilities and imposing liability—or attributing guilt—on the basis of an aggregation of probabilities results in the imposition of liability—or in conviction—of a defendant even if it cannot be established (under the relevant standard of proof) what misdeed he committed. In tort, failure to prove precisely *all* the detailed facts concerning the wrongful act does not preclude the attribution of liability. Thus, “if a car parked at the curb by the defendant begins to roll downhill” and the reason for this could be that the defendant “either failed to set the brakes or failed to cut the wheels properly against the curb, or failed to put the car in parking gear,” the trier of fact will find the defendant liable even without knowing exactly why he was at fault.²¹ But this case diverges from the criminal cases we focus on in this article: whereas in the latter, the indeterminacy relates to completely different misdeeds, in the former, it relates to different components of the same misdeed (the wrongful parking of a car).

The tort law doctrine that comes closest to the APP is the doctrine of Market Share Liability (“MSL”). This doctrine was applied by some courts in the DES cases. A drug designed to prevent miscarriages, DES, was manufactured by hundreds of companies mainly in the 1950s and turned out to be latently carcinogenic to female fetuses. Twenty-five years later, many of the young women whose mothers had taken the drug were diagnosed with uterine cancer. It was found by the courts that the drug had

²¹ For this example and others, see DOBBS, *supra* note 19, § 154, at 372.

not been tested adequately prior to its marketing and that the manufacturers had failed to take into account certain findings that had pointed to a risk of carcinogenic effect. Furthermore, the plaintiffs' mothers had never been cautioned against this risk. Finally, the drug had been marketed under a generic rather than brand name, which foiled attempts to trace each pill back to its actual manufacturer.²² For the purpose of providing a remedy to the victims, the courts developed the MSL doctrine. Under this doctrine, first adopted by the California Supreme Court in *Sindell*,²³ every defendant was to be held liable for the plaintiff's damage unless he could successfully prove that he did not manufacture the drug taken by the plaintiff's mother. As the *Sindell* court further clarified in its decision, this liability would be imposed only on those manufacturers who had produced a substantial proportion of the DES drugs in the relevant market. The court ultimately decided that the burden of compensating each plaintiff for her damage would be allocated amongst the manufacturers in accordance with their respective shares of the DES market.²⁴

The MSL is, in fact, a doctrine that amounts to an aggregation of probabilities in the judicial decision-making process in a way that resembles the APP. To better understand why, imagine that there are ten manufacturers in the market who wrongfully and separately produced and sold an identical hazardous product (like DES) to consumers, thereby causing identical injuries to a thousand people. Assume also that all the manufacturers have identical shares in the market and that it is completely impossible to trace any injury back to any specific manufacturer. In a single case brought by a single plaintiff, the probability that any single manufacturer caused the injury is .1, which is far below the required threshold for liability. However, the probability that a single manufacturer caused at least .1 of the total harm, namely, the sum of harms caused to all victims, is high and more than sufficient to justify imposing liability on that manufacturer. The MSL leads to this very result: once all suits have been resolved each manufacturer will bear .1 of the total harm. The MSL is, therefore, an analogous civil principle to the APP criminal principle: both aggregate probabilities and determine liability accordingly.²⁵ It should be

²² See *Sindell v. Abbott*, 607 P.2d 924 (Cal. 1980).

²³ *Id.*; *Collins v. Eli Lilly*, 342 N.W. 2d 37 (Wis. 1984), *cert. denied*, 469 U.S. 826 (1984) (applying MSL); *Martin v. Abbott Lab.*, 689 P.2d 368 (Wash. 1984) (same); *Hymovitz v. Eli Lilly Co.*, 539 N.E.2d 1069 (N.Y. 1989) (same). On the rejection of the MSL in Ohio law, see *Kurcz v. Eli Lilly & Co.*, 113 F.3d 1426 (6th Cir. 1997).

²⁴ It is not clear whether this decision should be interpreted as imposing liability on each defendant for all the plaintiff's damage (and then the proper allocation would be achieved through indemnification claims between the co-defendants) or as imposing liability on each defendant for only part of the damage. See *Brown v. Superior Court*, 751 P.2d 470, 485-87 (Cal. 1988) (adopting the second interpretation).

²⁵ Both principles differ from the alternative liability principle set by the California Supreme Court in *Summers v. Tice*, 199 P.2d 1 (Cal. 1948), which bears some superficial resemblance to the APP. In *Summers*, the defendants were two individuals who had participated in quail

noted, however, that the MSL has been applied almost only in cases involving identical conduct and identical risks created by all wrongdoers toward all the victims. Most of the courts that were willing to apply the MSL to DES cases refused to apply it in the absence of these features.²⁶ The corresponding criminal cases would, thus, be those in which the criminal acts attributed to the defendants with various probabilities are identical. At the same time, the MSL was applied to the DES cases even though there were numerous victims and the probability of a given single defendant having caused injury to a given specific victim was rather small. Hence the MSL is premised on the view that defendants can be found liable even when no specific harm to a specific plaintiff can be attributed to them.

Great caution must be taken when expanding criminal liability on the basis of analogies with tort law. Tort law and criminal law have significantly different goals, and the doctrines in each field should be responsive to those goals. Aggregating probabilities could serve deterrence, and not surprisingly, in torts, the main justification for the MSL is the concern to provide potential tortfeasors with efficient incentives.²⁷ Deterrence is also held to be an important goal of criminal law. Yet unlike tort law, retributive and expressivist considerations play a central role in this field.²⁸ This could explain why a more compelling case for aggregating probabilities can be made in tort law than in criminal law. In Parts III and V, we will return to deterrence, retribution, and expressivism.

hunting. The plaintiff had been shot in the eye by a stray bullet negligently fired by one of the defendants. The defendants had pulled their triggers simultaneously, so it could not be determined whose bullet had actually injured the plaintiff. The court resolved the case by establishing the “alternative liability” principle, which shifts the burden of proof from the plaintiff to the defendant “to absolve himself, if he can.” *Id.* at 86. Defendants unable to evidentially disassociate themselves from the damage are, thus, held liable for the damage in its entirety. This principle ultimately found its way into the Restatement (Second) of Torts § 433B, III. 9 (1965), but has nothing in common with the aggregation of probabilities being dealt with in this article. In *Summers*, there was a 50% probability for each of the defendants that he had hit the plaintiff, and this probability was not the result of any aggregation. It seems that the only aggregation of probabilities that could be conducted would be on the plaintiff—rather than defendant—side: the probability that the plaintiff suffered an injury from wrongful shooting would be the aggregate of the probabilities that each defendant had separately caused the injury. This would yield a probability of 1.

²⁶ For some exceptions, see ARIEL PORAT & ALEX STEIN, TORT LIABILITY UNDER UNCERTAINTY 60-67 (2001) (discussing cases in which the MSL was applied).

²⁷ See *id.* at 130-59 (discussing the justifications for the MSL). See also Mark A. Giestfeld, *The Doctrinal Unity of Alternative Liability and Market-Share Liability*, 155 U. PA. L. REV. 447, 481 (2007) (explaining that “[u]nder this method of apportionment, the interest of the DES plaintiff who has established a right to receive compensation for the injury from the group of defendants exactly corresponds to the interest of each individual defendant as a member of the causal group”).

²⁸ For a different view, see Ronen Perry, *The Role of Retributive Justice in the Common Law of Torts: A Descriptive Theory*, 73 TENN. L. REV. 177 (2006) (arguing that retributive justice has a certain influence on the development of tort law doctrines).

B. Pattern of Behavior and Interdependence

At a first glance, one could confuse the APP with the Prior Acts and Similar Crimes Doctrines. But in fact, these two doctrines, which we call "the pattern of behavior doctrines," work differently from the APP and rest on completely different grounds. Whereas the pattern of behavior doctrines are based on the interdependence of the offenses attributed to the defendant, the APP is most appropriately applied when those offenses are entirely independent of one another. Complete interdependence of the offenses, however, is not necessarily a reason not to apply the APP; under certain conditions, the Principle would apply even when they are interdependent, regardless of whether the pattern of behavior doctrines apply.

This section presents the pattern of behavior doctrines, comparing them with the APP and explaining under what circumstances interdependence would limit the application of the APP.

1. Prior Acts and Similar Crimes Doctrines

Under the Prior Acts Doctrine, which was adopted in Rule 404(b) of the Federal Rules of Evidence,²⁹ the prosecution can bring evidence of other crimes or acts that can be attributed to the defendant in order to establish motive, opportunity, intent, preparation, plan, knowledge, identity, or absence of mistake or accident. This evidence cannot be used to prove the defendant's bad character, and courts are required to instruct the jury accordingly.³⁰ Interestingly, under Rule 404(b) as interpreted by the

²⁹ Rule 404(b) prescribes that evidence of other crimes, wrongs, or acts is not admissible to prove the character of a person in order to show action in conformity therewith. It may, however, be admissible for other purposes, such as proof of motive, opportunity, intent, preparation, plan, knowledge, identity, or absence of mistake or accident, provided that upon request by the accused, the prosecution in a criminal case shall provide reasonable notice in advance of trial, or during trial if the court excuses pretrial notice on good cause shown, of the general nature of any such evidence it intends to introduce at trial.

See also *United States v. Woods*, 484 F.2d 127 (4th Cir. 1973), which was decided before the enactment of Rule 404(b) and where the court ruled,

Unlike other cases where evidence of prior crimes is admissible for only limited purposes and where it is necessary or proper to give limiting instructions, evidence of the prior events was admissible here to prove both that Paul was the victim of infanticide and that defendant was the perpetrator of the crime.

Id. at 137.

³⁰ See also *People v. Quinn*, 194 Mich. App. 250, 486 N.W.2d 139 (1992) (chastising the prosecution for arguing propensity based on evidence admitted under Rule 404(b)). A federal court asked to admit other bad acts under Rule 404(b) need not reach a preliminary finding that the defendant committed those acts. Rather, it need only determine that there is sufficient evidence to support a finding by a jury that the defendant committed the act submitted under Rule 404(b). See *Huddleston v. United States*, 485 U.S. 681 (1988), where

Supreme Court, even conduct that has been the subject of a prior acquittal can be submitted as evidence by the prosecution in a subsequent trial in order to support conviction.³¹ As Judge Easterbrook of the 7th Circuit has interpreted it in a concurring decision, the Rule is not a one of admissibility as it "says that evidence 'may' be admissible for a given purpose, not that it is automatically admissible."³²

The Similar Crimes doctrine, adopted in Rules 413 and 414 of the Federal Rules of Evidence, applies to sexual assault and child molestation offenses. Under this doctrine, if the defendant is accused of one of these types of offenses, "evidence of the defendant's commission of another offense or offenses of sexual assault or child molestation is admissible, and may be considered for its bearing on any other matter to which it is relevant."³³

The superficial similarity between the pattern of behavior doctrines and the APP stems from their shared feature: all three consider the past behavior of the defendant, and that behavior influences the likelihood of conviction.³⁴ But this resemblance notwithstanding, there is a substantial difference between them. The pattern of behavior doctrines are rooted in the premise that a person who has committed several offenses in the past is more likely to either have intended to commit or to actually have committed the offense he is presently accused of. His past behavior thus modifies the probability of his guilt in the current case. It is the *interdependence* between the past offense and the present alleged offense that provides the grounds for conviction. In contrast, the APP is based on the claim that when there is a specific probability that a person committed offense A and a specific probability that he committed offense B, the

the petitioner challenged his conviction for possessing stolen property on grounds that the trial court had improperly admitted evidence of "similar acts" involving his previous sale of stolen television sets. The judgment was affirmed on appeal, with the Supreme Court ruling that the trial court had not been required to make a preliminary finding that the petitioner had proved commission of the similar acts by a preponderance of the evidence. Evidence of other crimes is usually submitted in criminal, not civil, procedures. Rule 404(b), however, contains no such limitation, and potential civil applications occasionally arise. *See Barnes v. City of Cincinnati*, 401 F.3d 729 (6th Cir. 2005) (in a suit filed by a police officer against the City of Cincinnati, the court ruled that a statement made by a high-ranking official regarding lesbians in the city's police department was admissible under Rule 404(b)).

³¹ *See Dowling v. United States*, 493 U.S. 342 (1990) (testimony according to which the defendant had committed a crime, which had been brought in a prior trial that ended in acquittal, was decided to be rightly admitted under Rule 404(b) by the court in a subsequent trial, as establishing the defendant's identity and consequent guilt of bank robbery).

³² *U.S. v. Jones*, 455 F.3d 800 (7th Cir. 2006).

³³ Under Rule 415 of the Federal Rules of Evidence, this doctrine is applicable also to civil cases involving sexual assault and child molestation. For a critique of these rules, see Louis M. Natali Jr. & R. Stephen Stigall, "Are You Going to Arraign His Whole Life?": *How Sexual Propensity Evidence Violates the Due Process Clause*, 28 LOY. U. CHI. L.J. 1, 29 (1997) ("by requiring the admission of propensity evidence, the rules prevent a fundamentally fair trial, and thus violate due process").

³⁴ As Example 2 illustrates, sometimes the APP leads to acquittal rather than to conviction.

probability that he committed at least one of the offenses is greater than the probability that he committed A or the probability that he committed B. The APP is, therefore, in no sense based on any interdependence between the offences attributed to the defendant: the probability that he committed one offense does not change the probability that he committed another offense; only the probability that he committed *an unspecified offense* is effected.³⁵ Indeed, as we show next, sometimes interdependence is in fact a reason *not* to apply the APP.³⁶

2. Interdependence

To illustrate the difference between the APP and the pattern of behavior doctrines and to understand under what circumstances interdependence precludes the use of the APP, let us return to Example 1 and its defendant, who is being tried for two unrelated offenses. In this scenario, it is quite obvious that the patterns of behavior doctrines are not applicable, while the APP is. But assume now that the two offenses are sex offenses, say, sexual assaults. In these circumstances, the Prior Acts and Similar Crimes Doctrines could be applied to bring evidence of prior acts to support the allegation that the defendant either committed the sexual assaults or had the required intention to do so. The evidence relating to each of the two charges would then bolster the case against the defendant with respect to the other charge.³⁷ Indeed, a defendant who committed a

³⁵ The Racketeer Influenced and Corrupt Organization Act ("RICO" Act) can be interpreted as a tool for punishing individuals for unspecified offences. Under RICO, a person who is a member of an enterprise that has committed any two of specified crimes within a ten-year period can be charged with racketeering. Those found guilty of racketeering can be fined up to \$25,000 and/or sentenced to twenty years in prison per racketeering count. They are also subject to other civil sanctions. U.S. Code, Title 18, Part I, chap. 96. The racketeering offense can thus be seen as a mechanism for punishing individuals who are more likely to have committed serious unknown crimes. Arguably one can infer from the type of criminal activity committed by those convicted under RICO their engagement in other activities—activities that have not been proven. Yet it is quite evident that this is not the central purpose of RICO. The Act targets not those who are more likely to have committed other crimes but people whose criminal activity is particularly harmful because it contributes to organized crime. Hence, RICO cannot be construed as serving goals similar to that of the APP.

³⁶ Note that under the Prior Acts and Similar Crimes doctrines, the fact that a person committed several similar offenses in the past increases the chances of him being convicted in the present case. Under the APP, in contrast, as illustrated by Example 2, the fact that a person was convicted of committing several offenses in the past decreases the probability of his conviction in a later case. Unlike the Prior Acts and Similar Crimes doctrines, the APP dictates that the more a person has been convicted in the past, the higher the threshold required for a future conviction, and vice versa. But for reasons to be explored below, we do not suggest applying the APP across different trials. *Infra* Section IV.D.

³⁷ See *Gastineau v. Fleet Mortgage Corp.*, 137 F.3d 490, 495 (7th Cir. 1998) (admitting evidence of plaintiff's prior lawsuits to show, *inter alia*, "Gastineau's *modus operandi* of creating fraudulent documents in anticipation of litigation against his employers").

sexual assault in the past is more likely to have committed a later act of sexual assault.

Assume, however, that even with the application of the Prior Acts and Similar Crimes doctrines, none of the charges can be proven beyond reasonable doubt. Suppose that for each of the charges in the modified version of Example 1, examined in isolation, there is a probability of .7 that the defendant committed the offense, but once the two doctrines are applied, this probability increases to .9 for each. In the absence of the APP, the defendant would be acquitted of both charges; applying the APP would change the outcome to conviction. But, as we show below, it is unclear whether the APP would be applicable in such a case.

One central consideration in determining whether the APP should apply or not relates to the type of doubts the court has with respect to the defendant's guilt: if the same doubt exists with respect to all charges brought against the defendant the APP should not apply. In contrast, if there are different and independent doubts with respect to each offense, the APP should apply, either supplementing the pattern of behavior doctrines or as an alternative to them.

Example 3, another variation of Example 1, is illustrative of same doubt cases.

Example 3. Same Doubt. A person is charged with two offenses of sexual assault allegedly committed by him at different times and places with two different victims. When each case is examined separately, the evidence suggests that the probability that he committed each one of the offenses is .7. Applying the pattern of behavior doctrines increases this probability to .9. The reason the court is not fully persuaded that the defendant committed each of the offenses is *that it suspects that a specific person—the defendant's enemy—has framed him*. Assume that the required probability necessary to satisfy the beyond reasonable doubt standard is .95. Should the court convict the defendant for any of the offenses?

The answer is no: the APP should not apply and the defendant should be acquitted of both charges. If the defendant's enemy framed him in one case, it is likely that he framed him also in another. Therefore, there is a probability of .9 that the defendant committed the two offenses and a probability of .1 that he committed no offense at all. The probability that he committed only one offense is zero (or close to zero). Consequently, the interdependence between the two offenses attributed to the defendant precludes conviction.

It is also possible, however, that the doubts with respect to each of the charges to differ and is independent of one another, as the following Example demonstrates:

Example 4. Differing Independent Doubts. A person is charged with two offenses of sexual assault allegedly committed by him at different times and places against two different victims. The evidence in each case examined

separately indicates a probability of .7 that the defendant is guilty of each alleged offense. Applying the pattern of behavior doctrines increases that probability to .9. The reason the court is not fully persuaded that the defendant committed each one of the offenses is that, although it is clear that there was aggression on the part of the defendant in both cases, *it is not clear that the victims did not give their consent*. The absence of victim consent is a precondition for convicting the defendant under prevailing law. Assume that the required probability for satisfying the beyond reasonable doubt standard in the legal system is .95. Should the court convict the defendant of any one of the offenses?

In contrast to the case in Example 3, the answer here is yes, the APP should be applied. Since the court's doubts with respect to each charge are independent of one another (the victim in each case is different and so the respective doubts relating to each victim's consent are independent of one another), the probability that the defendant committed at least one of the sexual assault offenses is .99. In this case, then, the interdependence between the two charges, due to a pattern of behavior indicating a disposition to carry out the offenses, does not hinder applying the APP in tandem with the pattern of behavior doctrines.

In other cases, however, even if the doubt in each case were different, applying the pattern of behavior doctrines would preclude the use of the APP. If the pattern of behavior doctrines are not applied, however, the APP could be applied. Example 5 illustrates such cases.

Example 5. Differing Interdependent Doubts. A person is charged with two offenses of sexual assault allegedly committed by him at different times and places against two different victims. Examined separately, the evidence in each case indicates that a probability of .7 that the defendant committed each one of the offenses. When the pattern of behavior doctrines are applied, this probability increases to .9. The reason the court is not persuaded that the defendant committed each the offenses is *that in each cases there is a different lone eyewitness whose reliability is questionable*. Assume that the probability required for the beyond reasonable doubt standard in the legal system is .95. Should the court convict the defendant of any of the offenses?

In this Example, once the pattern of behavior doctrines are applied, the APP should not. The reason for this is that in this situation even though theoretically possible, using the APP becomes too complicated. For if one eyewitness in the Example were in fact a liar (or had made a mistake), then not only would acquittal of the relevant offense be justified, but the probability of the defendant's guilt of the other offense would decrease (from .9 to .7). Thus even though aggregating the probabilities in such cases is theoretically possible, it is impractical. This, however, does not rule out using the APP instead of the pattern of behavior doctrines. In Example 5, the Principle would yield a probability of .91 that the defendant

committed at least one of the alleged offenses.³⁸ This by itself would not be grounds for conviction, but a different outcome would obtain if, instead of .7, the probability of the defendant's guilt in each charge when examined separately were to come to .8.³⁹

The following conclusions can be drawn from the above discussion:

1. The APP should not be applied when identical doubts exist with respect to all of the alleged offenses. This holds regardless of whether or not the pattern of behavior doctrines are applied (Example 3).
2. The APP should not be applied when the offenses raise differing doubts if the pattern of behavior doctrines are applied and the probability of the defendant's guilt on each charge after applying these doctrines depends on his guilt of the other alleged offenses (Example 5).
3. The APP should be applied when a different doubt arises with respect to each offense and the pattern of behavior doctrines are not applied (Example 1 and the variation of Example 4, where the pattern of behavior doctrines are not applied).
4. The APP should be applied when the doubts differ with respect to each offense, even if the pattern of behavior doctrines are applied, as long as the respective probabilities of the defendant's guilt in each offense after the doctrines have been applied, are not impacted by whether he is guilty of the other offenses (Example 4).

These conclusions reveal the very broad scope of cases in which it is appropriate to use the APP. As long as the pattern of behavior doctrines are not applied—and note that they have quite narrow application under prevailing law—interdependence of the offences becomes irrelevant as long as the doubt with respect to each charge has a different source. Thus, in Example 1, even if we could show that the proportion of rapists among pickpockets were higher than in the general population, so that some degree of interdependence between the two offenses were to exist, it would not affect the suitability of the APP to the circumstances of this Example.

Furthermore, even when the pattern of behavior doctrines are applied, there are many situations—like that described in Example 4—that meet the condition of a lack of interdependence between the different charges after the pattern of behavior doctrines have been applied. Other examples of situations of this type are those in which the defendant's guilt in the offenses he is charged with rests on an outcome that is beyond his control. For example, suppose a defendant who is a gang member is charged with two murders, allegedly committed by him in different times and at different places in the presence of other gang members. Assume now that the court, after applying the Prior Acts doctrine, is sufficiently convinced that the defendant shot the victims in both cases with the requisite intention for committing murder, but in each case, the probability is .1 that someone

³⁸ $1 - (.3)^2 = .91$.

³⁹ $1 - (.2)^2 = .96$.

else's bullet hit the victim instead. Assuming again that .95 probability is required for conviction, the APP would result in a conviction for one murder and one attempted murder.⁴⁰

III. THE CASE FOR THE AGGREGATE PROBABILITIES PRINCIPLE

A. Adjudication Errors

Adopting the APP in criminal law can be expected to increase the number of errors in convicting the innocent (false positives or Type 1 errors) and decrease the number of errors in acquitting the guilty (false negatives or Type 2 errors). Later we explain why the APP does not *necessarily* increase error in convicting the innocent: in fact it even has the potential to reduce such error.⁴¹ It is beyond the scope of this article, however, to examine comprehensively the optimal mix of the two types of errors in criminal cases.⁴² Let us assume here, as we did before, that the “beyond reasonable doubt” standard implies a probability of at least .95 that the defendant committed the crime. Under this assumption, the law

⁴⁰ Bentham argued that when there is evidence that the same convicted person escaped detection by the law in the past, the sanction to be inflicted in the present conviction should reflect this fact. Bentham maintained that in setting the punishment, “it may be necessary, in some cases to take into account the profit not only of the individual offense to which the punishment is to be annexed, but also of such other offenses of the same sort as the offender is likely to have already committed without detection.” JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION, THE COLLECTED WORKS OF JEREMY BENTHAM 170 (J.H. Burns & H.L.A. Hart eds., 1996). We thank Avraham Tabbach for referring us to Bentham's thoughts on this issue. One way to interpret Bentham's argument is as the converse to our understanding of the Prior Acts and Similar Crimes doctrines: whereas in the latter doctrines, the court infers from past behavior forward to the present charge, Bentham urged courts to infer from the present charge backward to past behavior. The ramifications of this reading of Bentham's claim are that we can increase punishment in a present conviction in order to punish the convicted defendant for past behavior that, in light of the present conviction, can be more easily attributed to him now. Indeed, both the APP and Bentham's proposal are motivated by a concern for the under-enforcement of the law: the APP would be rendered completely meaningless if there were no under-enforcement and it had been always possible to fully and accurately detect all criminals. But as already explained, the APP is based on the conjecture of independence of the relevant probabilities, whereas Bentham's proposal is founded on the opposite assumption, namely, that if the defendant committed one offense, it is more likely he had committed other offenses in the past.

⁴¹ *Infra* text accompanying note 44.

⁴² See MITCHEL POLINSKY & STEPHEN SHAVELL, HANDBOOK OF LAW AND ECONOMICS 427-29 (2007) (discussing different ways to optimize type I and type II errors in law enforcement); I.P.L. Png, *Optimal Subsidies and Damages in the Presence of Judicial Error*, 6 INT'L REV. L. & ECON. 101-05 (1986) (same); Stein, *supra* note 9, at 141-71 (discussing the allocation of risks of error in the law of evidence).

prefers setting free 18—and not 19—guilty people rather than sending one innocent person to jail.

Suppose, now, that a person is accused of four offenses of similar severity and for each, there is a probability of .9 that he is guilty as charged. Adhering to the .95 threshold would mandate convicting the defendant for two offenses.⁴³ If the legal system acquits the person of all four offenses (as required by the DPP) it will seem to be endorsing a principle under which it is better to have 9,999 guilty people acquitted than one innocent person convicted. Aside from its evident absurdity, this outcome highlights the discriminatory effect of the DPP as opposed to the APP: upholding the DPP implies an unfair preference of people accused of committing a series of offenses as opposed to those accused of a single offense. The probability of guilt required to convict a defendant in the former case is much higher than what is necessary in the latter case, for the presumption of innocence applies in a differential and discriminatory manner. A person who has committed an offense beyond reasonable doubt will be set free simply because the particular offense that he committed could not be proven beyond reasonable doubt. The APP infers that the presumption of innocence should apply to the matter of whether the perpetrator has committed *an* offense rather than to the question of whether a particular offense was committed by that person.

As stated at the outset of this section, the APP can be expected to raise the number of erroneous convictions of the innocent. But interestingly enough, given certain realistic assumptions concerning the limited resources allocated to law enforcement, it is possible that the *costs* resulting from convicting the innocent would be lower under the APP than under the DPP and, in fact, even the *number* of errors resulting from convicting the innocent might be lower. Suppose there is a constraint on the total amount of punishment the state can inflict on offenders, for example, on the total number of years *all* offenders can be sent to prison. Since the APP is expected to yield more convictions, the state can take one of two strategies (or a combination thereof): *The one strategy* would be to shorten the period of time an offender is sent to jail for conviction on one charge. As a result, offenders who are convicted of several offenses would on average get more years in prison than what they currently get, and other offenders—those who are charged with having committed one offense—would on average get fewer years in prison than what they currently get. If the probability of error in a finding of guilt with respect to the former category of defendants is lower than with respect to defendants from the latter category—which is very likely—and since the costs of error in convicting the innocent is also a function of the years the innocent spends in jail, the shift from the DPP to

⁴³ *Supra* note 4.

the APP *could* decrease the total costs of convicting the innocent.⁴⁴ A *second strategy* would be to raise the minimum threshold necessary for conviction, for example from .95 to .98. With this strategy, the state could keep both the number of convictions and the average period of time an offender is sent to jail for one offense at its current level under the DPP. This strategy would decrease—rather than increase—the number of errors in convicting the innocent, since the required probability of guilt for conviction would be higher under the APP than is currently the case under the DPP.

B. Deterrence

The APP is superior to the DPP on deterrence grounds, particularly in its positive effect on the deterrence of repeat offenders. Under the APP, repeat offenders have a lower chance of avoiding conviction than under the DPP. This implies a higher expected sanction for repeat offenders under the APP and, accordingly, greater deterrence than under the DPP.

This advantage of the APP is especially significant if we assume that the expected sanction necessary to achieve optimal deterrence is in fact higher for repeat offenders than for other offenders.⁴⁵ But the APP is of

⁴⁴ Cf. Talia Fisher, Probabilistic Punishment (unpublished manuscript, on file with authors) (arguing that sanctions should be correlated with the probability of guilt and pointing out that, among other things, adopting such a rule could reduce the costs of convicting the innocent); Henrik Lando, *The Size of the Sanction Should Depend on the Weight of the Evidence*, 1 REV. L. & ECON. 278 (2005) (suggesting that sanctions be correlated with the weight of evidence and noting that this would result in less unfairness to the innocent who are wrongly convicted). Our argument is analogous to a different argument made by theorists, according to which it is justified to punish repeat offenders more severely than other offenders because the risk of wrongly convicting the innocent is lower with the former than with the latter. See C.Y. Cyrus Chu, Sheng-cheng Hu, Ting-yuan Huang, *Punishing Repeat Offenders More Severely*, 20 INT'L REV. L. & ECON. 127 (2000) (arguing that increasing the punishment for repeat offenders and decreasing it for other offenders could achieve the same level of deterrence and, at the same time, would reduce the risks of convicting the innocent); Richard Posner, ECONOMIC ANALYSIS OF LAW 227-33 (6th ed. 2003) (increasing punishment for repeat offenders is justified because the risk of convicting the innocent is lower in their case).

⁴⁵ There are different views on the question as to whether, in order to achieve optimal deterrence, repeat offenders should be punished more severely than other offenders. See Richard A. Posner, *An Economic Theory of the Criminal Law*, 85 COLUM. L. REV. 1215 (1985) (“a repeat offender is usually punished more severely than a first offender even if the repeat offender served in full whatever sentences were imposed for the earlier crimes”); David A. Dana, *Rethinking the Puzzle of Escalating Penalties for Repeat Offenders*, 110 YALE L.J. 733 (2001) (arguing that declining penalties for repeat offenders are optimal since the probability of detection escalates with offense history); Mitchell Polinsky & Daniel Rubinfeld, *A Model of Optimal Fines for Repeat Offenders*, 46 J. PUB. ECON. 291 (1991) (claiming that when the penalty is a fine and when the ill-gotten gains of the offenders are not considered part of the social good, it is optimal to punish repeat offenders more severely than other offenders in one type of cases, less severely in another type of cases, and with the same severity in other types of cases); Mitchell Polinsky & Stephen Shavell, *On Offense*

greater consequence from a deterrence perspective when we account for the fact that repeat offenders are often (although certainly not always) “professionals,” whereas one-time offenders are often (although certainly not always) amateurs.⁴⁶ As professionals, repeat offenders are likely to be more sophisticated and calculating than their first time counterparts; they are more responsive to sanctions and also more inclined to take “precautions” (or avoidance measures) to reduce the likelihood of conviction.⁴⁷ Consequently, repeat offenders, especially the most sophisticated ones, aware of the operation of the DPP, may seek to organize their criminal activity in such a way that foils sufficient evidence being amassed with respect to each distinct crime. Heads of crime organizations are a good example of such repeat offenders. Indeed, due to the DPP, many of them are not even brought to trial, let alone convicted. They are extremely proficient at playing by “the rules of the game” and the DPP facilitates this. It is this class of criminals that will be better deterred under the APP.⁴⁸

History and the Theory of Deterrence, 18 INT’L REV. L. & ECON. 305 (1998) (arguing that when the ill-gotten gains of the offenders are considered part of the social good, it is optimal to punish repeat offenders more severely than other offenders); Ariel Rubinstein, *On an Anomaly of the Deterrent Effect of Punishment*, 6 ECON. LETTERS 89 (1980) (arguing that punishing repeat offenders more harshly increases deterrence of offenders); Thomas J. Miceli & Catherine Bucci, *A Simple Theory of Increasing Penalties for Repeat Offenders*, 1 REV. L. & ECON. 71 (2005) (claiming that repeat offenders should be punished more severely than other offenders, because of their diminished employment opportunities); Winand Emons, *A Note on the Optimal Punishment for Repeat Offenders*, 23 INT’L REV. L. & ECON. 253 (2003) (arguing that when punishment is fine, under certain conditions, the optimal sanction scheme decreases); Winand Emons, *Escalating Penalties for Repeat Offenders*, 27 INT’L REV. L. & ECON. 170 (2007) (arguing that under certain conditions, increasing sanctions for repeat offenders is optimal and, under other conditions, the reverse holds.).

⁴⁶ See, for example, the “dangerous special offender” statute, 18 U.S.C. 35753 (1970), which provides for an enhanced penalty of up to twenty-five-years imprisonment for repeat offenders, professional criminals, and organized crime offenders.

⁴⁷ For a discussion of optimal enforcement when some individuals are more sophisticated than others, see Lucian Bebchuk & Louis Kaplow, *Optimal Sanctions and Differences in Individuals’ Likelihood of Avoiding Detection*, 13 INT’L REV. L. & ECON. 13 (1993).

⁴⁸ A possible counter-argument, applied to some of the cases, is that repeat sophisticated offenders may increase their avoidance efforts under the APP, which would be of greater benefit to them than under the DPP. Under certain conditions, this would result in more, rather than less, crime. See JACOB NUSSIM & AVRAHAM TABBACH, *DETERRENCE AND AVOIDANCE* (2005), available at <http://ssrn.com/abstract=844828> (showing that under certain conditions, higher sanctions encourage criminals to take more avoidance measures and reduce their expected sanctions). See also Chris William Sanchirico, *Character Evidence and the Object of Trial*, 101 COLUM. L. REV. 1227 (2001) (arguing that if bad character evidence were admitted at the conviction stage, the disincentive for engaging in crime would be weakened, since character evidence enhances the probability of conviction, both for those who committed the prescribed acts and for those who refrained from such behavior, leading to a decrease in the marginal cost of engaging in the criminal activity ex ante; banning bad character evidence thus promotes deterrence).

Deterrence considerations would not favor the use of the APP across different trials in cases characterized by Example 2 (when the APP leads to fewer, rather than more, convictions). The reason is that applying the DPP in such cases could result in the absurd outcome that a person who was convicted in the past is “free” (or, at least freer) to commit a crime with no punishment. To illustrate, if a person was convicted in the past for committing an offense and the probability of his guilt was established to be .95 (assuming, again, that .95 is the minimum threshold probability necessary for conviction), under the APP, he would be acquitted of any future offense so long as the probability of his guilt is established to be less than 1. For this and other reasons,⁴⁹ we do not recommend using the APP across different trials.

C. Costs of Enforcement

Another advantage of the APP is its cost-effectiveness. This feature results from the fact that the marginal costs of gathering items of evidence to prove a single specified offense increase. To illustrate, suppose that under the DPP, the prosecution needs to provide ten items of evidence to meet the standard of proof for a specific offense. It is typically much harder—and costly—to collect the tenth item of evidence than the ninth item, the eighth item, and so on.⁵⁰ Under the APP, nine items could be more than enough to secure a conviction, so long as the prosecution can provide one or more items of evidence relating to other offenses reasonably attributed to the defendant.

The premise that the marginal costs of evidence (brought to convict a person of a single offense) increase is not always true, however. Occasionally, collecting the tenth item of evidence of a single specified offense can be less costly than collecting the first item for a different offense. But this does not undermine our claim, since the APP would present the prosecution with an option: to either collect the tenth item of

⁴⁹ See *infra* Section IV.D.

⁵⁰ See, e.g., Chris William Sanchirico, *Evidence Tampering*, 53 DUKE L.J. 1215, 1333 (2004) (“The phenomenon of increasing marginal costs corresponds to the exhaustion of economies of scale in enforcement.”); Talia Fisher, *The Boundaries of Plea Bargaining: Negotiating the Standard of Proof*, 97 J. CRIM. L. & CRIMINOLOGY 943, 950-51 (2007) (“One can posit a situation where the task of proving the final X percent of the prosecution’s case requires a vast investment in resources on its part The prosecution may regard this evidence as crucial for proving its case ‘beyond reasonable doubt’”); Ehud Guttel & Alon Harel, *Matching Probabilities: The Behavioral Law and Economics of Repeated Behavior*, 72 U. CHI. L. REV. 1197, 1213 (2005) (stating that increasing enforcement can be achieved by either requiring enforcers to work more or recruiting additional personnel, under both of which “the marginal cost of enforcement is likely to increase”); Richard J. Gilbert & Michael L. Katz, *When Good Value Chains Go Bad: The Economics of Indirect Liability for Copyright Infringement*, 52 HASTINGS L.J. 961, 970 (2001) (“If it is relatively easy to detect some infringers, but not others, this pattern may lead to decreasing returns to scale (i.e., increasing marginal costs of enforcement at a given stage)”).

evidence for the original offense or else collect the first item (or more) for another offense. The prosecution presumably would make efficient use of this option and choose the less costly course of action. Under the DPP, no such option exists.

IV. PRACTICAL OBJECTIONS

In this Part, we discuss several objections to the APP. These objections reflect genuine concerns that should be taken seriously. Yet, as shortly to be shown, none of the objections provides a sufficient reason to reject the APP altogether. Instead, awareness of the force of these objections and the resulting modifications of our proposal can be used to refine the optimal scope of the rule.

We ignore, however, one practical constraint. Rule 8(a) of the Federal Rules of Criminal Procedure provides that two offenses may be joined in the same indictment if they “are of the same or similar character, or are based on the same act or transaction, or are connected with or constitute parts of a common scheme or plan.” Under Rule 14(a), courts may order separate trials if the joinder of offenses appears to prejudice a defendant or the government. Hence, it appears that much of what we propose is currently precluded by the prevailing procedural rules. We assume that if our proposal is adopted, some of these rules may be need to be changed.

A. Manipulations by the Prosecution and Agency Costs

Arguably, the use of the APP will inevitably invite large-scale abuse.⁵¹ After all, it is relatively easy to bring *some* evidence indicating the guilt of

⁵¹ For the risk of prosecutorial misconduct as a consideration in shaping procedural and evidentiary doctrines, see Angela J. Davis, *The American Prosecutor: Independence, Power and the Threat of Tyranny*, 86 IOWA L. REV. 393, 408-15 (2001) (asserting that prosecutorial misconduct occurs at numerous stages of the criminal process—including the pretrial stage and during trial—and that only on rare occasions is the misconduct discovered); Angela J. Davis, *Prosecution and Race: The Power and Privilege of Discretion*, 67 FORDHAM L. REV. 13, 20-25 (1999) (discussing prosecutors' vast discretion and power); Bruce A. Green, *Policing Federal Prosecutors: Do Too Many Regulators Produce Too Little Enforcement?*, 8 ST. THOMAS L. REV. 69, 70 (1995) (noting that “evidence of prosecutorial misconduct, particularly in federal cases, may be difficult to obtain”); Lesley E. Williams, *The Civil Regulation of Prosecutors*, 67 FORDHAM L. REV. 3441, 3442-47 (1999) (discussing how professional norms and statutory and constitutional law fail to regulate prosecutorial behavior in light of prosecutorial immunity); Charles P. Bubany & Frank F. Skillern, *Taming a Dragon: An Administrative Law for Prosecutorial Decision Making*, 13 AM. CRIM. L. REV. 473, 476-77 (1976) (noting the lack of controls over prosecutorial decision-making). Prosecutorial misconduct was one of the main concerns expressed by Justice Brennan in his dissent in *Dowling v. United States*, 493 U.S. 342, 363 (1990) (“The court today adds a powerful new weapon to the Government’s arsenal Indeed there is no discernible limit to the Court’s rule; the defendant could be forced to relitigate these facts in trial after trial.”).

any defendant. Consequently, there is the possibility that under the APP, any person could be convicted for some offenses without any significant evidence pointing his specific guilt. The prosecution would find it very easy—too easy—to "tailor" charges and abuse the criminal process. If a person is accused of committing a certain offense and the prosecution fails to prove his guilt beyond reasonable doubt, it could easily collect *some* evidence suggesting, with a low degree of probability, that the defendant committed a different offense and thus overcome the evidential hurdles to conviction.

This objection is not a reason for rejecting the APP, but, rather, for designing it in a way that would maximally alleviate the manipulation concerns. Recall that under the principles of evidence law, a person cannot be convicted on the basis of statistical evidence alone.⁵² The APP does not change the rules precluding conviction for statistical evidence.⁵³ Consequently, the APP would not allow the conviction of a driver for a violation of traffic regulations just because all drivers commit such violations on a daily basis.⁵⁴ In order to convict a person, case-specific evidence must be brought *with respect to each of the relevant charges*. Otherwise, conviction is not possible.

This may not fully alleviate this concern. In particular, it does not preclude the possibility of collecting low probability evidence for conviction. In order to prevent the abuse of the system with "tailored" accusations, it is possible to adopt "a minimum threshold" for case-specific evidence. Thus, for example, a standard could be set whereby only a probability above .5 that the defendant committed the offenses attributed to him (i.e., the preponderance of evidence standard) can be aggregated and used against the defendant. A minimal threshold of this type would significantly reduce the risks of abuse, which are particularly heightened when the threshold required for providing evidence is low.

A different objection to the APP relates to the concern that prosecutors would too often bring evidence that is insufficient for conviction for a specified offense but sufficient for conviction for an unspecified offense. This would happen—the argument goes—even when it is possible to get a conviction for a specified offense. If we assume that conviction for a

⁵² *Infra* Section IV.B.

⁵³ *Infra* Section IV.B.

⁵⁴ See Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts*, 98 HARV. L. REV. 1357, 1377 (1985) (discussing the Blue Bus Paradox, a hypothetical based on *Smith v. Rapid Transit, Inc.*, 317 Mass. 469, 58 N.E.2d 754 (Mass. 1945), where a person who is hit by a vehicle can only identify it as having been a bus, but sues the Blue Bus Company, proving that it owns 80% of the buses that run on the road where the accident occurred). Variations on the Blue Bus Paradox can be found, for example, in David Kaye, *The Limits of the Preponderance of the Evidence Standard: Justifiably Naked Statistical Evidence and Multiple Causation*, 1982 AM. BAR FOUND. RESEARCH J. 487, 488-89 (red and green taxi cabs).

specified, rather than unspecified, offense has added value,⁵⁵ this concern should weigh against adopting the APP.

We believe this objection to be unpersuasive. *First*, it is based on the premise that it is better to convict a person for a specified offense. Yet, at least in terms of efficiency-based considerations, this assumption is questionable. *Second*, even if we accept that convicting a defendant for a specified—rather than unspecified—offense has a greater value, for either justice-based or efficiency-based considerations, there is no reason to assume that prosecutors will ignore this added value when deciding on their strategy in building the case against a defendant. *Third*, if prosecutors are, regardless, expected to ignore the added value of convicting for a specified, rather than unspecified offense and, consequently, over-use the APP, this misuse of prosecutorial discretion should be tackled directly by the prosecutors' superiors. It is no reason to reject the APP.

Lastly, one could object to the APP because it could provide incentives for the prosecution to bring several charges of low probability against defendants rather than one high probability charge, not only to economize on enforcement costs, but also to create a heavier burden for the defendant to rebut the charges. Again, we think that this concern should be addressed directly by the courts as is often done in cases of abuse. For example, courts could reduce the sanctions imposed upon defendant when the prosecution brought charges which failed to satisfy even the preponderance of the evidence threshold. That would disincentivize the prosecution from bringing charges of low probabilities just in order to impose unjust burdens on the defendant.

B. *Statistical Evidence and Statistical Inferences*

Arguably, the APP is based on the use of what evidence law scholars call “statistical evidence,” evidence that is often regarded as inadmissible. Moreover, establishing the defendant's guilt under the APP is based on a probabilistic conception of the beyond reasonable doubt standard, which, according to some scholars, undermines the trust of the public in the criminal justice system.

Let us begin with the “statistical evidence” objection. In discussing the difference between “naked statistical evidence” and “trace-based forms of evidence,” Alex Stein writes,

Naked statistical evidence affiliates to the predictive, as opposed to the trace-based, mode of fact-finding. The predictive mode of fact-finding is invariably generalized. Fact finders endorsing this mode of reasoning assume that regularities observed in the past will reproduce themselves in future cases with roughly the same frequency as in the past. The trace-based mode—under which “proving that a nail was struck by a

⁵⁵ For example, see the law's expressivist functions at *infra* Part V.

hammer is to examine the head of the nail and there discover the trace of a hammer blow”—is case specific and individualized in character (because each trace is unique). Trace evidence, therefore, can always be tested for its connection to the individual defendant, which is not the case with predictive evidence.⁵⁶

The opposition to the use of naked statistical evidence is justified on the grounds that statistical inference “cannot be tested for its connection to the individual defendant.”⁵⁷ An individual defendant who is charged with fraud (or perhaps arson) simply because four houses owned by him were destroyed by fire within a relatively short period of time is arguably helpless against these charges since they are based exclusively on statistical inferences rather than on case-specific information.⁵⁸

It is beyond the scope of this article to discuss the pros and cons of using naked statistical evidence in criminal cases.⁵⁹ It is important, however, to establish that opposing naked statistical evidence, whose use is currently advocated by theorists (and perhaps adopted by courts), does not preclude the use of the APP.

Opponents of the APP could argue that the APP allows for inappropriate use of statistical evidence. A defendant convicted on the basis of the APP faces several charges, each of which was established with a certain probability. It is, therefore, statistical inference, rather than case-specific evidence, that led to the conviction. There is no connection between the inference leading to conviction and the particular circumstances of the defendant, since any defendant facing similar charges (and similar evidential support for these charges) would be convicted.

Yet, unlike the paradigmatic case of statistical evidence, the defendant is not helpless against the charges in the APP context since *ultimately* the charges rest on case-specific evidence. The defendant can rebut the charges simply by providing case-specific evidence concerning the separate individual offenses—evidence that will cast doubt on the probabilistic judgments. The APP rule does not therefore diverge fundamentally in this respect from the standard DPP rule in which a person is charged with a single well-specified offense and there are several separate items of evidence supporting the charge, each of which is not sufficient in itself for conviction but can, in aggregate, ground a conviction.

A conviction under the APP is typically based on the accumulation of all *case-specific* evidence brought against the defendant with respect to all

⁵⁶ STEIN, *supra* note 9, at 206-07.

⁵⁷ STEIN, *supra* note 9, at 206.

⁵⁸ This example is borrowed from STEIN, *supra* note 9, at 207.

⁵⁹ See, e.g., Nesson, *supra* note 54, at 1379 (stating that cases based only on probabilistic evidence are unlikely to reach the jury because “the fact finder cannot reach a conclusion that the public will accept as a statement about what happened”); Henry M. Hart Jr. & John T. McNaughton, *Evidence and Inference in the Law*, in EVIDENCE & INFERENCE 48, 54 (D. Lerner ed., 1958) (“[T]he law refuses to honor its own formula when the evidence is coldly ‘statistical.’”).

of the different charges and on the conclusion that the available evidence is sufficient to establish beyond reasonable doubt that the defendant committed at least one of the offenses. A comparison of two cases can demonstrate this point. In the one case, a certain amount of evidence is required for conviction under the DPP, e.g., 10 *case-specific* pieces of evidence. In the second case, a certain amount of evidence is required for conviction under the APP, e.g., 8 pieces of evidence relating to one offense and 8 pieces relating to a second offense. There is no difference in the types of evidence brought in the two cases. Since opponents of naked statistical evidence are willing to tolerate conviction in the former case, they also ought to tolerate conviction in the latter case.

Another objection to the APP—related to the statistical evidence objection discussed so far—is that the probabilistic conception of the beyond reasonable doubt standard, upon which the Principle rests, weakens public trust in the criminal justice system. The most prominent advocate of the public trust argument is Charles Nesson. According to Nesson, "the criminal justice system seeks to produce authoritative finality by inducing the general public to defer to jury verdicts."⁶⁰ In order to achieve this goal, "the evidence must do more than establish a statistical probability of the defendant's guilt: it must be sufficiently complex to prevent probabilistic quantification of guilt ... as long as the evidence prevents specific quantification of the degree of uncertainty, an outside observer has no reasonable choice but to defer to the jury's verdict."⁶¹

It is beyond the scope of this Article to discuss and evaluate Nesson's objection to probabilistic calculus in deciding criminal cases. For our purposes, suffice it to say that there are no grounds for any concern that applying the APP would undermine public trust in the criminal justice system. The Principle would enable jurors and judges to conclude that it has been proven beyond reasonable doubt that one, unspecified offence of several offenses was committed by the particular defendant, even if it cannot be established which offence specifically was committed. *Explicit* probability calculus is no more necessary for applying the APP than for the DPP when several pieces of evidence are being evaluated by the court.

C. Increased Litigation

Another objection to the APP is the concern that it would trigger a flood of litigation. This seems to be a natural consequence of the fact that the APP requires that courts consider even relatively low probability offenses in determining liability. This, the APP opponents may maintain, would encourage the prosecution to bring as much evidence as it can reasonably amass with respect to *any* seemingly criminal behavior on the

⁶⁰ Nesson, *supra* note 10, at 1195.

⁶¹ Nesson, *supra* note 10, at 1199.

part of the defendant in order to convince the court that the defendant is guilty of at least one offense. It seemingly follows, then, that the APP would generate a significant increase in the complexity of litigation and, accordingly lead to an increase in the costs of the criminal law system.

While the APP is undoubtedly likely to trigger more complex litigation, we believe this to be a trivial concern, for two reasons. *First*, adopting a minimal threshold that precludes courts from aggregating low probabilities offenses, as we suggested above, would mitigate the expected increase in litigation. *Second*, if the volume of litigation becomes too high due to the APP, it would be more sensible to decrease it by increasing the threshold for conviction than to reject the APP.⁶² *Third*, although the APP would stimulate more litigation, it is likely that the litigation costs for *one conviction* would decrease under the APP. As explained in Part III, under the APP, the costs of collecting evidence for any single conviction would be lower on average than under the DPP.⁶³ A similar rationale applies to litigation: the litigation costs entailed in increasing the probability of the defendant's guilt from .9 to .95 (as required under the DPP) can be expected to be higher on average than the cost of increasing this probability from, say, .5 to .55 (as allowed under the APP).

Fourth, and most importantly, the litigation generated by the APP is not frivolous. On the contrary, this increase in litigation would result in a correlative increase in justified convictions and better enforcement of the law.

D. Implementation Difficulties and the Meaning of a Probabilistic Threshold

The APP might be difficult to apply in practice. Courts could make mistakes and apply it improperly. A major potential source of difficulty is the interdependence of the offenses with which the defendant charged. Given the possibility of interdependence—so the objection goes—the application of the APP becomes too hard, both for courts and, especially, jurors. Furthermore, sometimes the interdependence is hidden. In such circumstances, the APP would be used as though there were no interdependence, which could excessively boost the false conviction rate. Lastly, even where interdependence does not present any problem for applying the APP, aggregating the probabilities could be too difficult a task for courts and especially jurors.

We discussed earlier the interdependence problem.⁶⁴ We established that so long as the doubts differ with respect to each offense and the pattern

⁶² Cf. *supra* Section III.A.

⁶³ See *supra* Section III.C.

⁶⁴ See *supra* Section II.B.

of behavior doctrines are not applied, there would be no reason not to apply the APP, *regardless of the lack or presence of interdependence*. This would encompass a large range of cases (Example 1 is illustrative) that could easily be handled by the courts. Moreover, even when the pattern of behavior doctrines are applied, as long as there are different doubts attaching to all of the offenses and the respective probabilities of the defendant's guilt of each offense after the pattern of behavior doctrines have been applied are not interdependent (Example 4), there is no reason not to apply the APP. We believe that judges would have no difficulty deciding whether these conditions have been met in any given case and, accordingly, could instruct the jury on whether or not to apply the APP.

The second problem, that of hidden interdependence, which precludes the application of the APP, warrants greater attention. To illustrate, suppose a driver is caught five times for speeding by the same police radar, and that radar's average rate of error is .75. Assume further that the radar's rate of error is higher in the evenings than in the mornings, say, because it is calibrated every night, and that the rate of error is especially high with regard to bright color cars. Say our driver was caught on five occasions in the evening and his car is bright colored. Assuming the court is unaware of the radar's defects, aggregating the probabilities could create an overly high risk of false conviction. But this risk of hidden interdependence is not a conclusive reason for rejecting the APP. Rather, courts should be mindful of this risk and require sufficient evidence to disprove its existence before the APP can be applied. Furthermore, given such awareness of the problem of hidden interdependence, the APP, if adopted, would generate inquiry and information concerning any possible interdependence between offenses.

Even in the absence of interdependence, applying the APP presupposes the court's knowledge of the probabilities relating to each offense, although in fact, courts do not possess such knowledge. To be sure, under the DPP, courts should be able to judge whether the beyond reasonable doubt standard has been satisfied, and this determination has at least some probabilistic features.⁶⁵ But the courts are not required to ascribe *accurate* probabilities to their findings on this matter, as opposed to what the APP seemingly requires. Moreover, it is even more difficult to accurately apply the APP when done across trials. In such a case, the court in Trial B needs accurate information about the probability of the defendant's guilt in Trial A, whether he was convicted or acquitted. This places an unreasonable burden on both courts A and B.

This objection is persuasive so long as it relates to aggregating probabilities across different trials. Indeed, we don't advocate aggregating probabilities across different trials. But this objection does not justify rejecting the use of the APP across different offenses in the same trial.

⁶⁵ See *supra* discussion accompanying notes 8-9.

Indeed, a court using the APP should look also at the "complete picture," namely, at all charges against the defendant and not only at each offense separately. However, it is important to note that the APP does not require courts to ascribe precise probabilities to each offense.

To understand just what the APP does require of courts, we can compare it to the requirements under the DPP. The DPP requires of courts to examine whether there is sufficient evidence that the defendant committed offense A, or sufficient evidence that he committed offense B, or sufficient evidence that he committed offense C and convict him of one of these offenses only if there is sufficient evidence that that particular offense was committed. In contrast, under the APP, the court needs to address the additional question of whether there is sufficient evidence that the defendant committed one of the three charged offenses.⁶⁶ Thus, using the APP, a court could conclude in a certain trial that even though it cannot convict the defendant for committing any *specific* offense, it can convict him for committing one *indeterminate* offense (or more) because there is no reasonable doubt that he committed one offense (or more).

This counter-argument to the implementation objection shows that there is no meaningful difference between how the reasonable doubt concept is applied under the APP and under the DPP. While under the latter Principle, the court convicts the defendant when there is no reasonable doubt he committed the specific offense attributed to him, under the APP, the defendant is convicted when there is no reasonable doubt that he committed at least one offense among several with which he is charged.

E. Redundancy

Finally, it can be argued that the APP is already being used by courts implicitly, and thus there is no need to recognize it explicitly. Moreover, if courts are indeed, applying it implicitly, forcing them to apply it explicitly may result in double counting. Under this objection, when several charges are brought against a defendant—even if unrelated to one another—the

⁶⁶ A similar argument to the one discussed here is sometimes raised against the application of the Hand Formula in torts, which, arguably, requires courts to calculate expected damages and costs of precaution and then compare them with each other in order to determine whether the defendant was negligent or not. However, in order to implement the Hand Formula, it is sufficient that the court determine whether the marginal expected damages are higher or lower than the marginal costs of precautions, and it need not make any accurate calculation of those figures. See ROBERT COOTER & THOMAS ULEN, LAW & ECONOMICS 349, 351-52 (5th ed. 2007) ("The marginal Hand rule states that the injurer is negligent if the marginal cost of his or her precaution is less than the resulting marginal benefit To apply the Hand rule, the decision-maker must know whether a little more precaution costs more or less than the resulting reduction in expected accident costs."). See also Ariel Porat, *Offsetting Risks*, 106 MICH. L. REV. 243, 272-73 (2007) (explaining how probabilistic rules can be applied with rough, rather than accurate, information about probabilities).

judge and certainly the jury are influenced by the accumulation of charges and tend to convict more easily than if there were only one charge.⁶⁷

Obviously, the argument that courts implicitly aggregate probabilities across offenses is valid only when several offenses are charged at the same trial. When this is not the case, aggregating probabilities is certainly not done implicitly and should also not be done explicitly. But if the APP is a desirable mechanism, it provides a justification for trying several unrelated charges, even of different natures, against one defendant in the same trial. Of course, there are, and should be, other considerations concerning this matter, which could have greater weight, but the desirability of the APP should also be regarded as a relevant factor.⁶⁸

Is it true, though, that courts implicitly aggregate probabilities across offenses? It is hard to know whether this is empirically right or wrong. At least with respect to judges, it may be possible to assume a certain commitment on their part to examine each charge separately and not to be influenced by the multiplicity of charges. The prevailing legal ethos founded on the DPP principle does not allow considerations of the type examined above. To the extent that judges inculcate this ethos, it follows that they are likely to consciously reject the very possibility of aggregating probabilities.

But if courts do sometimes apply a rule that resembles the APP, it is better that this be done explicitly and systematically, rather than implicitly and randomly. Furthermore, the application of the APP can sometimes be complicated and tricky, and it would be best to contend with this in a straightforward manner, rather than leaving it to the rough and, at times, inconsistent intuition of judges and jurors.

Arguably, even if judges do not apply the APP, it is possible that the police and prosecution in fact apply some version of the Principle in making their decisions regarding law enforcement efforts and bringing defendants to trial. According to this argument, when the police and prosecution acquire evidence related to different offenses allegedly committed by the same person, they are more likely to bring him to trial, they generally have more information about his potential involvement in perpetrating crimes, and more importantly, they try harder to collect more evidence, thereby increasing the chances of conviction.⁶⁹

⁶⁷ The Prior Acts and Similar Crimes doctrines allow courts and jurors, under certain conditions, to consider the accumulation of the evidence of all charges. But as we explained at *supra* Section II.B.1, these two doctrines differ from the APP.

⁶⁸ See *supra* Part IV (second paragraph). Interestingly, those who oppose aggregating probabilities across different offenses, both explicitly and implicitly, could make use of exactly the opposite argument: different charges should not be brought at the same trial to avoid the risk of aggregation of this type.

⁶⁹ David A. Dana, *Rethinking the Puzzle of Escalating Penalties for Repeat Offenders*, 110 YALE L.J. 733, 742-43 ("The question whether probabilities of detection escalate is ultimately an empirical matter, but not a matter easily subject to study. Because offenders are reluctant to provide candid information regarding their *undetected* violations, researchers

We find this argument to be unpersuasive. If courts refuse to apply the APP, it certainly will be a factor in prosecution decisions *not* to charge suspects even if, when aggregating the probabilities, they are convinced that a given person is guilty of a specific offense. Prosecutors will know that as long as they are unable to establish that the defendant committed a specific offense, the court will acquit him under the DPP. *Second*, even if the police and prosecution do increase their enforcement efforts vis-à-vis a person against whom there is enough evidence for conviction under the APP but not under the DPP, it is still not clear why the courts should not apply the APP. By refusing to adopt the APP, courts encourage the prosecution to incur excessive enforcement costs.⁷⁰ Furthermore, often, the police and prosecution will not succeed in bringing sufficient evidence for conviction under the DPP, even if, under the APP, conviction is warranted, either because of the prohibitively high cost of collecting more evidence or because the necessary evidence is impossible to collect.

Finally, it could be argued that, at least in plea bargains, the APP is already applied in practice: when there are several accusations against the defendant, and even when none of them meets the threshold necessary for conviction, the aggregation of the probabilities will influence the deal made between the prosecution and defendant. But even assuming this to be right, we do not see why courts should not apply the APP. As described above, the prosecution acts in the shadow of the prospective trial. Therefore, if the APP is not applied by courts, this will most certainly affect the shape of plea bargains, though *not* in the right direction. Furthermore, even if the APP is *perfectly* applied in the context of plea bargains, there is still no reason why the Principle should not be applied by courts as well.

V. RETRIBUTIVIST AND EXPRESSIVIST THEORIES OF PUNISHMENT

A. *The Case against Aggregation of Probabilities*

Thus far, we have established that deterrence-based theories, particularly theories that focus on efficiency, would be likely to endorse a moderate version of the APP. We now turn to consider the justifiability of the APP and DPP from the perspective of justice-based theories. The upshot of this Part is that different justice-based theories of punishment are likely to endorse different views of the APP. Some justice-based theories, in particular some versions of retributivist theories, would be inclined to accept the APP, while others, in particular expressivist theories, would tend to reject it.

face huge obstacles in developing any comparative assessments of the success of different groups of offenders in evading detection.").

⁷⁰ See *supra* Section III.C.

Let us proceed with retributivist theories. Kant maintained that “[p]unishment by a court ... can never be inflicted merely as a means to promote some other good for the criminal himself or for civil society. It must always be inflicted upon him only *because he has committed a crime*.”⁷¹ This observation lies at the foundation of many retributivist theories.⁷² Dolinko conceives retributivists as those who explain “either the rational justification of punishment, or its moral justification, or both, by appealing to the notion that criminals deserve punishment rather than to the consequentialist claim that punishing offenders yields better results than not punishing them.”⁷³ Under what Dolinko labels “bold retributivism,” “lawbreakers deserve punishment and that this, all by itself, constitutes a good or sufficient reason for the state to inflict punishment on them.”⁷⁴ Furthermore, asserts Dolinko, retributivists believe also in proportionality, namely, that wrongdoers ought to be made to suffer in proportion to their offenses.⁷⁵ Criminals, according to this view, simply deserve to be punished, and this desert provides the justification for inflicting punishment on them.⁷⁶ Hence, under this version of retributivism, if we establish that an agent committed a wrong, we have a reason to impose a sanction on that person, even if the nature of the wrong remains unspecified, since the person clearly deserves to be punished. It should be noted that there is no consensus as to what retributivism really is, but for the purposes of the discussion in this section, we focus on Dolinko’s version.

In contrast, expressivists underscore the importance of the expressive, educational, and communicative aspects of the criminal sanction. Under expressivist theories, sanctioning a wrongdoer is a public manifestation of condemnation and disapprobation of his deeds. Some believe that the need for condemnation is in itself sufficient justification for the infliction of criminal sanction, whereas others hold that it is conducive to other goals, such as education or the inducement of a sense of guilt.

Robert Nozick falls into the former camp. He believes that “[r]etributive punishment is an act of communicative behavior.”⁷⁷ In elaborating on the concept of communicative behavior, Nozick speaks of retributive principles as encompassing two aspects. The first is to “connect the wrongdoer to value qua value,” and the second is to connect the wrongdoer in a way “that the value qua value has a significant effect in [the

⁷¹ IMMANUEL KANT, *THE METAPHYSICS OF MORALS* 6:331 (Mary Gregor ed., Cambridge University Press 1996).

⁷² We do not argue, however, that Kant was committed to the versions of retributivism that we examine below. Kant’s theory of punishment has been examined by many theorists, and we do not purport to provide an accurate depiction of it here.

⁷³ David Dolinko, *Some Thoughts about Retributivism*, 101 *ETHICS* 537, 541-42 (1991).

⁷⁴ *Id.* at 542.

⁷⁵ See, e.g., Thomas E. Hill, *Kant on Wrongdoing, Desert and Punishment*, 18 *LAW & PHILOSOPHY* 407, 409 (1999).

⁷⁶ *Id.* at 425.

⁷⁷ ROBERT NOZICK, *PHILOSOPHICAL EXPLANATIONS* 370 (1981).

criminal's] life, as significant as his own flouting of correct values.”⁷⁸ Similarly, Joel Feinberg asserts that “punishment is a conventional device for the expression of attitudes, resentment and indignation, and of judgments of disapproval and reprobation, on the part either of the punishing authority himself or of those ‘in whose name’ the punishment is inflicted.”⁷⁹ Jean Hampton shifts the focus to educational concerns. In her view, “punishment is intended as a way of teaching the wrongdoer that the action she did (or wants to do) is forbidden because it is morally wrong and should not be done for that reason.”⁸⁰

Although expressivist theories do not have to necessarily reject the APP, they would seem likely to have reservations with respect to its applicability. After all, these theories highlight the condemnation or disapproval of *an act*, and a prerequisite for conveying condemnation and disapproval of an act is the identification of the object of condemnation and disapproval, i.e., the unambiguous identification of the act being condemned.⁸¹ Punishing a person for *an* offense she may or may not have committed (simply because it is highly probable that she committed either this offense or a more serious one) rather than for *the* offense she committed dilutes the important expressive, educational, and communicative message of punishment. Hence, expressivist theories would likely reject the APP because, under the Principle, no specific act can be attributed to the individual being punished and, consequently, no act can be effectively condemned. To condemn a person for an act that he may have not committed simply because the act is part of a disjunction of acts diverges significantly from condemning a specific act. It is only the latter that can meet expressivist concerns.

Thus, the rejection of the APP by the expressivist theorists is no accident. It is a byproduct of the way these theories address what seems to be one of their apparent weakness. Expressivist theories are always vulnerable to the accusation that condemning theft, rape, or murder does not necessitate the infliction of sanctions on the perpetrator.⁸² After all, these acts can be effectively condemned without any resort to punishment. To address this objection, expressivist theories claim punishment to be a special mode of expression. And the distinct nature of punishment as an expressive practice requires that the object of condemnation be specific and concrete. Punishment must, therefore, be designed to express disapproval

⁷⁸ *Id.* at 376.

⁷⁹ Joel Feinberg, *The Expressive Function of Punishment*, in *DOING & DESERVING* 95, 98 (1970).

⁸⁰ Jean Hampton, *The Moral Education Theory of Punishment*, 13 *PHILOSOPHY & PUB. AFFAIRS* 208, 212 (1984).

⁸¹ It is possible, of course, to develop an expressivist theory that focuses on the condemnation of the character of the actor or his culpability rather than condemnation of the acts he has performed. This is not the route taken by traditional expressivist theories of punishment.

⁸² See, e.g., Feinberg, *supra* note 67, at 87; Hampton, *supra* note 68, at 161.

of a particular act that was performed by the given perpetrator, not of an evil act that may or may not have been committed by him. Evidence of this requirement for specificity is abundant in the expressivist theories. According to Hampton, for instance, the punisher needs “to communicate to the wrongdoer that *her* victims suffered ... so that the wrongdoer can appreciate the harmfulness of *her* action.”⁸³ Feinberg also maintains that “punishment surely expresses the community’s strong *disapproval of what the criminal did.*”⁸⁴ Communicating disapproval by punishing an individual for a disjunction of acts does not satisfy the specificity of expression requirement of these theories.

Proponents of the APP could counter the expressivist claim with the argument that by endorsing the APP, punishment conveys disapproval of *all offenses* comprising the disjunction. Arguably, a conviction based on the APP can reflect disapproval and condemnation of *all* offenses included in the disjunction. Thus, ironically, it seems that the APP is an even more effective means of expressing disapproval than the DPP because it conveys the message that *all* the offenses included in the disjunction warrant condemnation.

This contention, however, fails to appreciate the subtlety of the concerns raised by expressivist theories. It fails to capture the significance of the condemnation of a concrete act—the precise act that has been perpetrated by the criminal. Concrete condemnation stresses the hideousness of an actual act performed by the defendant: murder, rape, theft, or fraud rather than merely a crime deserving a sentence of at least ten years such as murder *or* rape *or* theft *or* fraud or even the condemnation of both robbery and murder when only one can be attributed to the perpetrator.

Another reason why expressivist concerns may lead to rejecting the APP is the well-being of victims of crime. Victims often wish for the criminal who perpetrated the crime against them to be punished *for that crime*. But an implication of the APP is that the criminal cannot be convicted of any specific crime and, consequently, no victim of a particular crime can establish that a wrong has been committed against *him*. Hence, arguably, punishment for an unspecified crime does not provide victims of crime with recognition of the commission of a wrong against them.

To sum up, different justice-based theories take different stances with regard to the APP. Whereas at least some retributivist theories are likely to be sympathetic to the Principle, expressivist theories are likely to be more wary of it. Perhaps this explains the intuitive reluctance on the part of criminal law theorists and practitioners to implement the APP in practice.

⁸³ Hampton, *supra* note 68, at 162.

⁸⁴ Feinberg, *supra* note 67, at 76.

B. The Case for a Moderate Aggregate Probabilities Principle

When, then, if ever, should the APP be recognized and give rise to a conviction? For retributivists (at least advocating the type of retributivism described above), the answer would be that the APP should be applied without limit. In contrast, expressivist theories would give a more reserved response: it all depends on whether using the APP will effectively serve the expressive, educational, and communicative functions of criminal law. We argue that the more similar two offenses are, the more likely that applying the APP will not undercut the expressive, educational, and communicative functions attributed by expressivists to criminal law. In contrast, the greater the heterogeneity of the offenses, the greater the willingness of these theorists to apply the DPP is.

Similarity and difference are, of course, complex and multifaceted concepts. It is not always a simple feat to determine what makes two offenses similar or different in the relevant sense.

One parameter is the *nature of the offense*. To understand the relevance of the nature of the offense, assume that it can be proven beyond reasonable doubt that a person committed an act of either murder or theft but it cannot be established that he committed any one of them. The fact that the offenses are so different and that one is classified a bodily offense whereas the other is classified as a property offense seems sufficient to arrive at the conclusion that this person ought not to be convicted. The concern in this case has been acknowledged by some criminal law theorists as the concern of “fair labeling.” This is the concern that “offenders ... be labeled with an adequate degree of precision, in order that the criminal record identifies the gist of ... [the offender’s] criminal wrongdoing.”⁸⁵

It would also seem implausible under expressivist theories to convict a person who committed one of two unrelated fraud offenses, for example. If the prosecution can prove that the defendant committed either fraud on one occasion or an unrelated act of fraud on another occasion, he should most likely be acquitted. This example illustrates the relevance of a second important dimension of expressivist theories: *the homogeneity of the different instances of the same offense*. In the present example, the two offenses are classified as fraud offenses; the very same criminal law provisions would be applied against the perpetrators of these offenses. But, despite this formal similarity, no two fraud offenses are identical in severity. The nature and severity of any concrete fraud offense are always colored by the particular circumstances of the case at hand: the sum of money involved, the identity of the victim, etc. Heterogeneity makes it more difficult to express concrete condemnation of the act performed by

⁸⁵ A.P. Simester & G.R. Sullivan, *On the Nature and Rationale of Property Offenses*, in *DEFINING CRIMES* 186 (R.A. Duff & Stuart Green eds., 2005).

the defendant since the disjunction of the offenses consists of very different acts.

Yet there are circumstances under which the heterogeneity of the circumstances should not bar conviction under the expressivist theories. Assume that two bank officers have committed a series of (unrelated) frauds against a single bank during the same time period. It can be proven beyond reasonable doubt that one officer committed a series of frauds against the bank and stole \$100,000, while the other officer committed a series of frauds against the bank and stole \$200,000; it cannot, however, be established who committed which series of frauds. It seems, in such a case, that the *similarity in circumstances* is sufficient to make the condemnation of both offenders concrete enough and to convey a clear and concrete disapproval of the behavior in question. Another parameter that seems to bear relevance in this case is the *identity of the victim of the offense*. If it can be established that several offenses were committed against one particular victim and, in addition, that the circumstances under which the offenses were committed were identical, then expressivist theories could endorse the use of the APP even if it is not possible to establish which exact offense was committed by the defendant.

Our investigation of justice-based theories is inconclusive. On the one hand, it seems that retributivist theories (of the type discussed above) would favor the APP, whereas expressivist theories would be reluctant to accept it. Yet even the latter need not reject the Principle outright. Rather, the APP should warrant rejection by expressivists only when its use would frustrate the expressivist goals of the criminal law. The more similar the crimes composing the disjunction are, the less reluctant expressivist theories should be to endorse the APP. The inconclusiveness of this Part is not coincidental. It reflects genuine conflicting sentiments characterizing the practice of criminal law. The reluctance of the legal system to endorse the APP suggests that expressivist concerns play an important role in the practice of criminal law.

CONCLUSION

This Article investigated a puzzle: why the APP has been unequivocally and universally rejected in criminal law. It is our claim that the reason is rooted in expressivist theories of punishment, which suggest that the sanction meted out to a convicted offender should reflect disapproval of a particular act and that the act needs to be identified so that the disapproval is sufficiently concrete. This concern can explain the greater appeal of the APP in contexts where deterrence seems to be the primary objective, such as regulatory offenses.

If we put aside expressivist concerns, the APP should be adopted by the legal system. But even then, there would still be practicalities that would limit the scope of the Principle's application. An appropriately modified

version of the APP, which takes into account the objections discussed in this paper, would promote deterrence, minimize adjudication error, and save enforcement costs. Given such modifications, the APP could be made safe from abuse and tailored to be consistent with justice-based theories of punishment, including the expressivist theories.

We therefore suggest that the APP be applied with great caution and awareness of the difficulties it can generate. *First*, the Principle should be applied only to charges brought in the same trial and not charges brought in different trials. Implementing the APP across trials could be quite difficult for courts and could also amount to double jeopardy. In addition, if the APP is adopted, the current rules of criminal procedure should be changed to allow joinder of unrelated offenses in the same trial.⁸⁶ *Second*, the APP should be applied primarily to regulatory offenses or homogeneous offenses in order to satisfy expressivist concerns. We do suggest, however, considering broader application, especially in contexts where the risk of repeat offenders escaping punishment under the DPP is high. *Third*, in applying the APP, particular attention should be paid to the interdependence problem. Such sensitivity would be especially imperative when the pattern of behavior doctrines are applied by the court. *Fourth*, the APP should be applied only to those cases where the probability of the defendant's guilt is higher than .5. This restriction would reduce the risk of abuse by the prosecution and, at the same time, would allow courts a familiar standard of proof they are well trained to implement.

This Article is ultimately the byproduct of an enigma. It is rooted in an observation that the practice of law seems to reject out of hand and categorically what simple and common-sense reason seems to emphatically endorse. While in general the practice of law is wiser than theorists tend to imagine, it may at times be prone to error in judgment. We believe that the rejection of the APP is one such rare case.

⁸⁶

See supra Part IV (second paragraph).