
Article

Aggregating Probabilities Across Cases: Criminal Responsibility for Unspecified Offenses

Alon Harel and Ariel Porat[†]

INTRODUCTION

Should a court convict a defendant for an unspecified offense if there is no reasonable doubt that he committed an offense, even though the prosecution cannot prove his guilt as to a particular offense beyond a reasonable doubt? Stated otherwise, is committing *an* offense sufficient for a conviction or must a prosecutor establish what this offense is to justify a

[†] Alon Harel is the Phillip P. Mizock & Estelle Mizock Professor in Administrative and Criminal Law at the Hebrew University of Jerusalem and a Visiting Professor of Law at the 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, Avraham Tabbach, and Doron Teichman. 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, the 2009 Annual Meeting of the American Law and Economics Association held at San Diego University, the workshops at Bar-Ilan University, Cornell University, Haifa University, the Interdisciplinary Center in Herzliyah, Ono Academic College, the Center for Rationality at the Hebrew University, Tel Aviv University, 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. Lastly, we thank Julie Kaster from the *Minnesota Law Review* Editorial Board for her editorial assistance. Copyright © 2009 by Alon Harel and Ariel Porat.

conviction? This Article contends that, under certain conditions, a prosecutor should not have to establish the particular offense committed by a defendant—proof that the defendant committed *an* offense should be sufficient.

Two distinct methods exist by which judges and juries could evaluate whether the standard of “beyond a reasonable doubt” is satisfied in cases where a defendant is charged with several offenses. Under the traditional “distinct probabilities principle” (DPP), when a defendant is charged with a number of offenses, the court examines each charge individually to decide whether the beyond-a-reasonable-doubt standard is satisfied *with regard to each charge*. Alternatively, the court could use the “aggregate probabilities principle” (APP) to examine all charges in aggregate and decide whether the standard is satisfied with respect to *at least one* charge.¹ Example 1 illustrates how the APP would work.

Example 1: A person is charged with pickpocketing and rape, two unrelated offenses allegedly committed by him at different times and places. 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-a-reasonable-doubt standard is .95.² Should the court convict the defendant on either of the offenses?

Weighed separately—that is, applying the DPP—the defendant in Example 1 will likely be acquitted of both offenses. Yet, there is a .99 probability³ that he committed at least one offense, which is higher than the .95 probability necessary for conviction in a criminal trial. Alternatively, the APP would convict him of one unspecified offense and impose on him *at*

1. Cf. Saul Levmore, *Conjunction and Aggregation*, 99 MICH. L. REV. 723, 729 (2001) (noting the absence of an aggregate-probabilities approach in criminal law, according to which probabilities are aggregated across the different *elements* of the same offense). Levmore does not raise, however, the question of aggregating probabilities across different *cases*, which this Article addresses. Cf. *id.*

2. This Article uses numerical probabilities for the sake of exposition. See *infra* notes 6, 17 and accompanying text.

3. 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 $1 - .01 = .99$.

minimum the sanction of the less severe of the two offenses, i.e., pickpocketing.⁴ Example 1 raises a straightforward dilemma: If, using the DPP, the Example 1 defendant is acquitted of all offenses, he will escape conviction despite the fact that the probability that he committed at least one offense (.99) is greater than the probability required for criminal conviction (which is, under our assumption, .95). Individuals are routinely convicted for committing a single offense on the basis of evidence that establishes guilt with a lower probability (.95) than the aggregate probability that arose from the evidence in Example 1 (.99, for an unspecified offense). Applying the APP to Example 1 would allow the prosecution to establish that the defendant committed beyond a reasonable doubt at least one of the two offenses with which he or she was charged. Arguably, it is not just or efficient that the Example 1 defendant is acquitted while, at the same time, a defendant charged with a single offense that can be proven at a lower probability (.95 under our initial assumption) is convicted.

To be sure, the probabilities in Example 1 (and also in the other examples which follow) are merely illustrative. Courts generally do not ascribe numerical probabilities to defendants' guilt.⁵ Instead, when a court determines whether the reason-

4. If the defendant had been charged with four offenses instead of two, this would yield a probability of .9999 that he had committed at least one offense. Applying the APP would guarantee conviction for two offenses since the probability that two offenses had been committed would be higher than the threshold required for conviction. This is the outcome of a binomial distribution. There are four events and in each one the defendant either committed the offense or did not (thus he either committed zero, one, two, three, or four offenses and the probability that one of these scenarios transpired is one). To calculate the probability that the defendant committed at least two of the four offenses we subtract from one the probability that he committed zero offenses or one offense. Since the probability that the defendant did not commit any offense is $(.1)^4 = .0001$, and the probability that he committed exactly one 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 three offenses; we multiply by four because the specific offense committed by the defendant could be any of the four offenses), the probability that he committed at least two offenses is $1 - .0001 - .0036 = .9963$. To calculate the probability that the defendant committed at least three offenses we add the probability that he committed four offenses to the probability that he committed three offenses. Since the probability that the defendant committed four offenses is $(.9)^4$ and the probability that he committed three offenses is $(.9)^3 * .1 * 4$, the probability that he committed at least three offenses comes to $.6561 + .2916 = .9477$.

5. See, e.g., *McCullough v. State*, 657 P.2d 1157, 1159 (Nev. 1983) ("The concept of reasonable doubt is inherently qualitative.").

able-doubt standard was satisfied in a particular case, it employs a rough intuitive judgment which inherently encompasses at least some probabilistic component.⁶ Therefore, to understand the dilemma in Example 1, it is enough to recognize the truism that when there is a likelihood that event A took place (committing pickpocketing) and there is also a likelihood that event B took place (committing rape), and those events are not fully interdependent on each other, the likelihood that at least one of those events (A or B) took place is higher than the likelihood that either A took place or that B took place.⁷

Example 1 illustrates how the APP can result in more convictions than the DPP. The APP can also result in fewer convictions, as is illustrated in Example 2.

Example 2: A person is charged with pickpocketing and rape, two unrelated offenses, allegedly committed by him in different times and places. 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-a-reasonable-doubt standard is .95. Should the court convict the defendant on both offenses?

If the offenses are examined independently under the DPP, then the defendant will be convicted on both charges. Yet, the probability that the defendant committed *both* offenses is only .9, which is lower than .95.⁸ In contrast, under the APP, the court would convict the defendant of only one offense, since the probability that he committed at least one offense is greater than .95 (it is .9975). Even though similar rationales support the application of the APP to cases represented by Examples 1 and 2, the focus of this Article is on cases where the APP would result in more, rather than fewer, convictions—namely, those cases represented by Example 1.

6. The .95 threshold has traditionally been used as an illustration in texts that interpret the principle of beyond a reasonable doubt in probabilistic terms. 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.”).

7. See discussion *infra* Part II.B.2.

8. $(.95)^2 = .9025$.

Surprisingly, the possibility of using the APP in criminal law has yet to be explored.⁹ Courts have never discussed or considered the principle nor does case law suggest that a prosecutor or a defense lawyer has ever suggested applying it. Lawyers as well as theorists seem to take it for granted that a person can only be convicted for committing a specific identifiable crime.¹⁰ The APP deserves serious attention. Both justice and efficiency considerations support applying the APP to a broad range of cases. This Article addresses the most powerful objections to applying the APP in prevailing criminal law and proposes that courts in certain instances adopt the APP. If the APP is adopted as advocated, the presumption of innocence currently applied with regard to the offense will be replaced with a presumption of innocence applied with regard to the accused. Thus today, under the DPP, a defendant will be presumed innocent¹¹ and acquitted if the prosecution cannot persuade the court that he committed a specific offense beyond a reasonable doubt.¹² In contrast, under the APP, the accused will be presumed innocent and acquitted only if the prosecution cannot show that no reasonable doubt exists that the accused committed *any* offense.

9. In other fields, however, the APP has been considered and discussed at length. For instance, legal theorists have proposed aggregating probabilities in civil cases, and, as such, that discussion will not be explored in this Article. *See, e.g.*, Levmore, *supra* note 1, at 724 (discussing aggregating probabilities mainly in tort cases). Furthermore, Frederick Schauer and Richard Zeckhauser proposed aggregating probabilities across cases outside the judicial context. Frederick Schauer & Richard Zeckhauser, *On the Degree of Confidence for Adverse Decisions*, 25 J. LEGAL STUD. 27, 41–51 (1996). Schauer & Zeckhauser 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. Schauer & Zeckhauser maintain, however, that such an argument is inapplicable to criminal proceedings. *Id.* “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.

10. *See, e.g.*, *Wicks v. Lockhart*, 569 F. Supp. 549, 565 n.18 (E.D. Ark. 1983) (noting that the prosecution must prove beyond a reasonable doubt each offense charged).

11. *See, e.g.*, *Estelle v. Williams*, 425 U.S. 501, 517 (1976) (“One of the essential due process safeguards that attends the accused at his trial is the benefit of the presumption of innocence . . .”).

12. *See, e.g.*, *Wicks*, 569 F. Supp. at 565 n.18 (“It is axiomatic that the government bears the burden of proving beyond a reasonable doubt *as to each offense charged* . . .”).

Part I introduces the APP, explores its potential scope, and discusses the applicability of the APP to different types of cases. Part II distinguishes the APP from other types of aggregations of probabilities conducted in both criminal and tort law, such as the market share liability, and prior-acts and similar-crimes doctrines. This Part also explores the rare instances when use of the APP should be precluded from the outset. Part III argues that the APP should be adopted because it does a better job than the DPP of fulfilling the overarching social goals of criminal law—deterrence, efficient law enforcement, and minimization of adjudication errors. Part IV raises several possible practical objections to the APP, most importantly its potential abuse by the police and prosecution and the difficulties in its implementation, but shows that these objections are not compelling enough to justify rejecting the APP outright. Finally, Part V analyzes expressivist theories as means of explaining why courts are precluded from adopting the APP, but argues that even expressivists should recognize circumstances where conviction of the defendant on the basis of the APP is justifiable.¹³

I. INTRODUCING THE AGGREGATE PROBABILITIES PRINCIPLE

To convict a person in a criminal trial the prosecution must prove the charges “beyond a reasonable doubt.”¹⁴ The rationale underlying this requirement as well as its precise meaning, are, of course, controversial.¹⁵ Yet, it is undisputed that the standard has an important probabilistic aspect to it¹⁶: the evidence

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

14. *In re Winship*, 397 U.S. 358, 364 (1970) (“[W]e explicitly hold that the Due Process Clause protects the accused against conviction except upon proof beyond a reasonable doubt of every fact necessary to constitute the crime with which he is charged.”).

15. See Jessica N. Cohen, *The Reasonable Doubt Jury Instruction: Giving Meaning to a Critical Concept*, 22 AM. J. CRIM. L. 677, 678 (1995) (“[B]ecause reasonable doubt is a term of art it should be defined for the jury.”); Thomas V. Mulrine, *Reasonable Doubt: How in the World Is It Defined?*, 12 AM. U. J. INT’L L. & POL’Y 195, 197–98, 210–25 (1997) (explaining various approaches to and definitions of reasonable doubt); Lawrence M. Solan, *Refocusing the Burden of Proof in Criminal Cases: Some Doubt About Reasonable Doubt*, 78 TEX. L. REV. 105, 105 (1999) (“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.”).

16. See ALEX STEIN, FOUNDATIONS OF EVIDENCE LAW 65 (2005) (“Adjudi-

that grounds a conviction in a criminal trial must establish that the defendant committed an offense with a high degree of probability.¹⁷ Criminal and evidence law implicitly assume that convicting a criminal for committing an offense requires prosecutors to meet the beyond-a-reasonable-doubt standard with respect to each particular offense.¹⁸ This principle, the DPP, has never been questioned.¹⁹

The APP challenges the main tenets of the DPP. The APP's basic idea is that a defendant should be convicted for an offense when it is certain, or almost certain, that he committed some offense, even if the exact offense cannot be established. For example, under the APP, the defendant in Example 1 would be convicted for one offense, as the probability that he committed *no* offense is .01.²⁰ Or, if Example 1 is modified to encompass four offenses, instead of two, then the defendant should be convicted of some offense because the probability that he committed *no* offense is one in ten thousand.²¹ To reduce the risk of excessive punishment, the APP would only inflict on the defendant in Example 1 the least severe sanction (the punishment for pickpocketing). The APP may sometimes harm, and at other times benefit, defendants. Example 2 illustrates that under the APP a defendant accused of two offenses where each offense—if examined separately—can be proven beyond a rea-

cative 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.”).

17. For those readers who are skeptical about mathematical calculations in the legal context, it is possible to consider 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 one he specifically committed? 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) (“Both 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.”).

18. See, e.g., *Wicks v. Lockhart*, 569 F. Supp. 549, 549 (E.D. Ark. 1983).

19. *Id.* at 565 n.18.

20. See *supra* Example 1.

21. See *supra* note 4.

sonable doubt should not be convicted for both offenses, but rather, only for one.²²

Theoretically, the APP is applicable in all cases where a defendant is charged with more than one offense. There are, however, some distinctions between different types of cases that have potential normative significance in shaping the APP. First, cases where a defendant is charged with identical offenses are distinguishable from cases where he is charged with different offenses (the nature-of-the-offense criterion). As demonstrated later, theories of punishment, such as expressivism, which ascribe great importance to the expressive function of the criminal law, may insist that the APP not be used in cases of different offenses but would tolerate it when the offenses are similar or identical.²³ For example, expressivists would be reluctant to convict a person charged with pickpocketing or rape of either offense when neither can be proven beyond a reasonable doubt, even if it is almost certain that the person committed at least one of them.²⁴

A second distinction that may be relevant in shaping the APP relates to the homogeneity of the offense. Even if the offenses are identical in type, it would be easier—for expressivists in particular—to accept a 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 pickpocketing or breaching the statutory speed limit are typically less products of circumstances.

A third distinction relates to the crime victim’s identity. In some cases, the relevant offenses are directed at the same victim, whereas in other cases, different victims are the targets of the different offenses (the same-victim criterion). Thus, there

22. This is not to say, however, that the APP is neutral overall with respect to defendants. There are not an identical number of expected convictions and expected acquittals resulting from an application of the APP. The transition from the DPP to the APP can be expected to bring about more convictions than acquittals, based on the observation that the APP, in taking into account all probabilities from .01 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. *See supra* Example 2.

23. *See* discussion *infra* Part V.

24. *Id.*

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, expressivists may find it more acceptable to convict a person for an unspecified offense if it can be proven beyond a reasonable doubt that an offense was committed against a single victim. The message conveyed by conviction may serve to affirm the grievance of the victim. 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 may be greater. For example, 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.²⁵

Fourth, the use of the APP seems more reasonable when used to convict individuals for regulatory rather than criminal offenses. Regulatory offenses are governed primarily by considerations of deterrence²⁶ and expressivist considerations (which may preclude the use of the APP) are less applicable with respect to these offenses (the regulatory-offense criterion).²⁷ For example, imagine a defendant being accused of violating the speed limit on two different occasions and that the prosecution could not establish his guilt beyond a reasonable doubt with respect to each of the occasions. However, if the prosecution could establish that he violated the speed limit on at least one occasion, then the opposition to applying the APP is expected to be much weaker than it is when applying the APP in cases such as those discussed in Example 1.²⁸

The fifth distinction relates to the difference between cases in which the defendant is charged with all offenses simultaneously and those cases where he is charged with a new offense after having been previously convicted or acquitted of other offenses (the same-trial criterion). Compare Example 1, where there are two simultaneously charged offenses and the evidence

25. As we elucidate later, interdependence could sometimes be a significant obstacle in employing the APP. See discussion *infra* Part II.B.2.

26. See Charles J. Walsh & Alissa Pyrich, *Corporate Compliance Programs as a Defense to Criminal Liability: Can a Corporation Save Its Soul?*, 47 RUTGERS L. REV. 605, 633 n.89 (1995).

27. See *infra* Part V.

28. See *supra* Example 1.

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, where 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 counterintuitively, 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.²⁹ Yet the case for applying the APP across different trials, rather than different charges within the same trial, is weaker because it undermines the finality of judicial decisions and faces substantial challenges in implementation.³⁰ We do not advocate this approach.

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 where 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 case³¹: Suppose a murder and a burglary were committed at the same time in two different places, and hidden cameras recorded both incidents. Unfortunately (for society and for law enforcement authorities), the perpetrators of these crimes are twin brothers. It is known, therefore, that

29. If the defendant was acquitted in the first trial because the probability of his guilt was only .9, and in the second trial the probability of his guilt was again .9, under the APP he should be convicted at the second trial. If, instead, that defendant was convicted at the first trial because the probability of his guilt was .95, under the APP he should not be convicted at the second trial. Moreover, if in the latter case the probability of guilt in the second trial was also .95 (and not .9), as illustrated by Example 2, applying the APP should also lead to acquittal at the second trial.

30. 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* discussion *infra* Part IV.D. Second, taking into account prior acquittals as a consideration for convicting the same defendant in a subsequent trial could violate the Double Jeopardy Clause. *See* U.S. CONST. amend. V. Third, applying the APP across different trials in cases represented by Example 2 (when the APP generates fewer, rather than more, convictions) could 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* discussion *infra* Part III.B.

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

each of the two brothers committed one of the offenses. It is unknown, however, which offense was committed by which brother. Under the APP, both brothers would be convicted for the lesser of the two crimes, namely, burglary.

These considerations establish that the APP may raise both theoretical and practical difficulties and it may not be possible to apply it across the board. We will later see that these difficulties do not undermine the importance of the APP, yet are relevant to determining its precise scope. The main conclusion that we reach at the end of the Article is that among the various distinctions discussed above, the most important one is the homogeneity criterion: the more heterogeneous the crimes are, the more that the expressivist concerns preclude the use of the APP. We suggest, however, that when the application of the APP is strongly supported by deterrence, the APP should apply even to heterogeneous offenses.

II. THE AGGREGATE PROBABILITIES PRINCIPLE IN CONTEXT

A. AGGREGATING PROBABILITIES UNDER PREVAILING LAW

Aggregating probabilities is not an unfamiliar phenomenon in the legal system.³² This Part is devoted to examining cases of aggregation which are already familiar to the legal system and showing how these cases differ from the APP.

The issue of aggregating probabilities takes place in fact-finding procedures when a judge or jury determines whether a conjunction of facts or events transpired.³³ For each of the facts or events comprising the set, there is a specific probability correlating to each fact or event. Aggregating these probabilities enables the judge or jury to determine whether the plaintiff or the prosecution proved a conjunction of *all* the facts or events to a sufficient degree.³⁴ Alternatively, the probability that at least one fact took place requires an aggregation of a different sort.³⁵

32. See, e.g., *People v. Collins*, 438 P.2d 33, 33, 40 (Cal. 1968); ROBERT COOTER & THOMAS ULEN, *LAW & ECONOMICS* 460–66 (5th ed. 2008) (discussing the aggregation of probabilities in tort law).

33. See Robert Cooter, *Adapt or Optimize? The Psychology and Economics of Rules of Evidence*, in *HEURISTICS AND THE LAW* 379, 380 (G. Gigerenzer & C. Engels eds., 2006).

34. *Id.* at 384–87.

35. See Levmore, *supra* note 1, at 729–30 n.11 (labeling this alternate method “reverse conjunction”).

Let us start with the first scenario. Suppose a judge in a civil case must decide (1) whether the defendant was negligent and (2) whether he caused the given injury. Only if both questions are answered affirmatively will the defendant be found liable.³⁶ Assume that the judge or jury estimates the probability of the defendant's negligence at .6 and the probability that he caused the injury at .6 as well. Aggregating the probabilities yields a probability of .36 that the defendant was both negligent *and* caused the injury (the civil-cumulative case). If the judge's or the jury's decision rests on the .36 aggregation, then the plaintiff will fail to satisfy the preponderance-of-the-evidence standard and the defendant will be found not liable.³⁷ If each component of the cause of action is considered separately, however, 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,³⁸ and case law seems inclined against using aggregation in such cases.³⁹

Similar difficulties arise in cases of disjunctive liability where the defendant is liable if either scenario A or scenario B took place. Assume that scenarios A and B each involve con-

36. See RESTATEMENT (SECOND) OF TORTS § 281(b)–(c) (1965) (listing these two elements as required for negligence liability).

37. See Alex Stein, *Of Two Wrongs That Make a Right: Two Paradoxes of the Evidence Law and Their Combined Economic Justification*, 79 TEX. L. REV. 1199, 1205 (2001) (explaining the rationale of aggregating probabilities in the civil-cumulative case).

38. To consider the applicability of the “product rule”—which is the rule that mandates the aggregation of probabilities—compare Maya Bar-Hillel, *Probabilistic Analysis in Legal Factfinding*, 56 ACTA PSYCHOLOGICA 267, 269 (1984) (“[T]he conjunction of a small number of weakly probative characteristics can be strongly probative.”), and Bernard Robertson & G. A. Vignaux, *Probability—The Logic of the Law*, 13 OXFORD J. LEGAL STUD. 457, 478 (1993) (“Once one regards probability as a generalisation 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.”), with L. JONATHAN COHEN, *THE PROBABLE AND THE PROVABLE* 58–67 (1977) (discussing the problems of conjunction of facts or probabilities and claiming that mathematical probability is inadequate as a model for rational thinking). See also Stein, *supra* note 37, at 1203–05 (considering the effect of the “conjunction paradox” on the use of the product rule, but suggesting that, in light of another major distortion in fact-finding, the product rule leads to a second-best solution).

39. See Levmore, *supra* note 1, at 752 nn.58–60 (arguing that no jurisdiction explicitly recognizes the product rule 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).

flicting types of negligence, either of which could have caused the injury suffered by the plaintiff. Assume that the probability of scenario A is .3, that 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 because the probability that at least one of the scenarios occurred amounts to .51.⁴⁰ Thus, aggregating the probabilities in such a case (the civil-alternative case) would generate a different decision than when each probability is considered separately.⁴¹ The legal system is ambivalent with respect to aggregation in such cases.⁴² As in the civil-cumulative case, courts generally reject the aggregation principle in civil-alternative cases.⁴³ Yet some important exceptions exist. In tort law, for instance, proving by a preponderance of the evidence that the defendant's wrongful act caused an injury is normally sufficient to impose liability, even if the plaintiff cannot prove by the same standard what *precisely* made the defendant's act wrongful.⁴⁴

Finally, aggregating probabilities is also relevant in criminal cases.⁴⁵ As discussed above with respect to civil cases, if several components of the offense must be proven to establish the defendant's guilt, then aggregating the probabilities of each

40. The probability that none of the events took place is $.7 * .7 = .49$. 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 $.3 + .3 = .6$.

41. See Levmore, *supra* note 1, at 729 n.11, 745–46 (explaining the “alternative routes” scenario, which he labels “reverse conjunction”). Levmore uses the same reasoning for rejecting the product rule in the civil-alternative case as in the civil-cumulative case. See *id.* at 752 nn.58–60.

42. Compare Tribe, *supra* note 17, at 1361 (claiming that hard statistical data lead decisionmakers into “[d]warfing [the] soft variables” by assuming that “[i]f you can't count it, it doesn't exist”), with Jonathan J. Koehler & Daniel N. Shaviro, *Veridical Verdicts: Increasing Verdict Accuracy Through the Use of Overtly Probabilistic Evidence and Methods*, 75 CORNELL L. REV. 247, 265 (1990) (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”).

43. See Levmore, *supra* note 1, at 729 n.11 (stating that courts do not apply the product rule).

44. DAN DOBBS, *THE LAW OF TORTS* § 154, at 370–73 (2001) (describing *res ipsa loquitur* cases where the jury is permitted to infer that the defendant was negligent in causing the harm in a specific scenario, even though evidence of any specific negligent act cannot be established).

45. See Bar-Hillel, *supra* note 38, at 268–70, 282–83 (analyzing the use of probabilities in cases and suggesting a “soft role . . . for probability in the fact-finding process”). *But see* *People v. Collins*, 438 P.2d 33, 33, 40 (Cal. 1968) (rejecting the use of probabilities in determining guilt on the facts of the case).

component separately generates a different outcome than an integrated aggregation of the probabilities.⁴⁶ For instance, if convicting a person for burglary requires both trespass and intent to commit a crime, the judgment could be different in accordance with the method of aggregation. Thus, even if each component of the offense (e.g., trespass and intent) can be proven beyond a reasonable doubt, reasonable doubt could still exist with respect to the cumulative presence of all components (the criminal-cumulative case).⁴⁷ Will the court convict the defendant under such circumstances? The answer is unclear.⁴⁸

The APP differs substantively from aggregation in both the civil and criminal cumulative cases because it does not question when courts ought to convict a defendant for a particular specific offense. Aggregating probabilities in civil and criminal cumulative cases determines whether the person committed the particular wrongful act or offense. Liability in a civil context or conviction in a criminal context implies that the court was satisfied that the evidence justified the imposition of liability or criminal conviction for a *particular* act. Conversely, a finding that a defendant is not liable (in a civil context) or not guilty (in a criminal context) implies that the court cannot justify liability or conviction for a *particular* act. This Article, however, examines alternative rather than cumulative cases: cases in which no specific offense can be attributed to the defendant although it is evident (or at least sufficiently probable) that the defendant committed an offense. Indeed, in contrast to the civil and criminal cumulative cases, the APP is about substantive criminal law rather than fact-finding. This Article focuses on the *commission of an unspecified offense* and does not discuss

46. See *supra* note 37 and accompanying text.

47. See Levmore, *supra* note 1, at 729 (suggesting that the product rule can equally apply to the civil-cumulative case and the criminal-cumulative case).

48. Compare *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), with 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’” (citation omitted)).

how to establish whether any particular offense has been committed.

The civil-alternative case is much more relevant to the APP. The APP involves aggregating the probabilities and imposing liability on the basis of that aggregation even if it cannot be established what misdeed the defendant committed. In tort law, a failure to prove precisely *all* the detailed facts concerning the wrongful act does not preclude the attribution of liability.⁴⁹ 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,” then the trier of fact could find the defendant liable even without knowing exactly why he was at fault.⁵⁰ But the issues raised by civil tort claims, such as this one, diverge from the criminal cases we focus on in this Article: whereas in the former, it relates to different possible misdeeds related to the same act (the wrongful parking of a car), in the latter, the indeterminacy relates to completely different acts.

The APP is closely related to the tort law doctrine of market share liability (MSL), which was applied by some courts in the diethylstilbesterol (DES) cases.⁵¹ 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.⁵² Twenty-five years later, many of the young women whose mothers had taken the drug were diagnosed with cancer of their reproductive organs.⁵³ Courts found that the drug had 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 effects.⁵⁴ Furthermore, the plaintiffs’ mothers had never been cautioned against this risk.⁵⁵ Finally, the drug was marketed under a generic name, which foiled plaintiffs’ attempts to trace each pill back to its actual manufacturer.⁵⁶ For the pur-

49. See *supra* note 44 and accompanying text.

50. For this example and others, see DOBBS, *supra* note 44, § 154, at 372.

51. See *Sindell v. Abbott Labs.*, 607 P.2d 924, 936–38 (Cal. 1980) (holding drug manufacturers liable even when proving which specific manufacturer produced the drug in question is not possible).

52. *Id.* at 925.

53. *Id.*

54. *Id.* at 925–26.

55. *Id.* at 926.

56. *Id.*

pose 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*,⁵⁷ every defendant properly joined would 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.⁵⁸ As the *Sindell* court further clarified in its decision, liability would be imposed only on those manufacturers who had produced a substantial proportion of the DES drugs in the relevant market.⁵⁹ The court ultimately decided that the burden of compensating each plaintiff for her damage would be allocated among the manufacturers in accordance with their respective shares of the DES market.⁶⁰

The MSL doctrine amounts to an aggregation of probabilities in the judicial decision-making process in a way that resembles the APP. To better understand how, imagine that there are ten manufacturers in the market who produced and sold an identical hazardous product (like DES) to consumers, thereby causing identical injuries to one 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 a 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—the sum of harms caused to all victims—is more than sufficient to justify imposing liability on that manufacturer. The MSL doctrine leads precisely to this result: once all suits have been resolved

57. *Id.* at 937; *see also* *Hymovitz v. Eli Lilly & Co.*, 539 N.E.2d 1069, 1078 (N.Y. 1989) (applying a modified version of the MSL doctrine); *Martin v. Abbott Labs.*, 689 P.2d 368, 381 (Wash. 1984) (same); *Collins v. Eli Lilly [&] Co.*, 342 N.W.2d 37, 49 (Wis. 1984) (same). *But see* *Kurzei v. Eli Lilly & Co.*, 113 F.3d 1426, 1433 (6th Cir. 1997) (rejecting application of the MSL doctrine under Ohio law).

58. *Sindell*, 607 P.2d at 937.

59. *Id.*

60. *Id.* It is not clear whether this decision should be interpreted as imposing liability on each defendant for all the plaintiffs' damage (with the proper allocation 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 of imposing several liability on each defendant for only part of the damage); ARIEL PORAT & ALEX STEIN, TORT LIABILITY UNDER UNCERTAINTY 138, 148 (2001) (discussing the policies behind the two alternative rules).

each manufacturer will bear .1 of the total harm as there is sufficient evidence that he caused at least .1 of the total harm.

The MSL doctrine is, therefore, an analogous tort law principle to the APP criminal law principle: both aggregate probabilities and determine liability accordingly.⁶¹ It should be noted, however, that the MSL doctrine has been applied almost exclusively in cases involving identical conduct and risks created by all wrongdoers toward all the victims.⁶² Most of the courts that applied the MSL doctrine to DES cases refused to apply it in the absence of these features.⁶³ The corresponding criminal cases would thus be those where the criminal acts attributed to the defendants with various probabilities are identical. At the same time, the MSL doctrine was applied to the DES cases even though there were numerous victims and the probability of a single defendant having caused injury to a specific victim was rather small.⁶⁴ Hence, the MSL doctrine is premised on the view that defendants can be found liable when no 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 different goals,⁶⁵ and the doctrines in each field

61. 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 hunting. *Id.* at 2. The plaintiff had been shot in the eye by a stray bullet negligently fired by one of the defendants. *Id.* The defendants pulled their triggers simultaneously, so it could not be determined whose bullet had actually injured the plaintiff. *Id.* 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 4. Thus, “[d]efendants unable to disassociate themselves evidentially from the damage are, therefore, held liable for the entire damage.” PORAT & STEIN, *supra* note 60, at 61. This principle ultimately found its way into the RESTATEMENT (SECOND) OF TORTS § 433B, illus. 9 (1965), but has nothing in common with the aggregation of probabilities dealt with in this article. In *Summers*, there was a fifty percent probability for each of the defendants that he had hit the plaintiff, and this probability was not the result of any aggregation. 199 P.2d at 2. It seems that the only aggregation of probabilities that could be conducted would be on the side of the plaintiff rather than the defendant: 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.

62. PORAT & STEIN, *supra* note 60, at 64–65 (discussing the limits of the MSL doctrine and citing cases).

63. *But see id.* at 65–67 (2001) (discussing cases in which the MSL doctrine was applied).

64. *See id.* at 60–62.

65. *See generally* OLIVER WENDELL HOLMES, JR., THE COMMON LAW 34–

should be responsive to those goals. Aggregating probabilities could serve as a deterrent and, not surprisingly, in torts the main justification for the MSL doctrine is to provide potential tortfeasors with efficient incentives.⁶⁶ Deterrence is also held to be an important goal of criminal law.⁶⁷ Yet unlike tort law, retributive and expressivist considerations play a central role as well.⁶⁸ This could explain why a more compelling case for aggregating probabilities can be made in tort law than in criminal law.

B. PATTERN OF BEHAVIOR AND INTERDEPENDENCE

This Section is divided into two subsections. First, it differentiates the APP from two existing doctrines: the prior-acts and similar-crimes doctrines.⁶⁹ Under both of these doctrines, past similar behavior on the part of the defendant can be used as evidence supporting conviction.⁷⁰ But these two doctrines, termed “the pattern-of-behavior doctrines,” are distinct from the APP. Whereas the pattern-of-behavior doctrines are based on the probabilistic interdependence of the offenses attributed to the defendant, the APP is most appropriately applied when those offenses are entirely independent of one another.

Second, this Section explores the relevance of the interdependence of the offenses to the application of the APP, and establishes the conditions under which interdependence precludes such application. Interdependence of the offenses, however, is not necessarily a reason not to apply the APP. As is shown, under certain conditions, the APP would apply even when offenses are interdependent, and occasionally even in tandem with the pattern-of-behavior doctrines.

43 (Transaction Publishers 2005) (1881) (comparing the different goals of the criminal law); PORAT & STEIN, *supra* note 60, at 1–15 (considering the evolution of Anglo-American tort doctrine).

66. See PORAT & STEIN, *supra* note 60, at 130–59 (discussing the justifications for the MSL doctrine); Mark A. Geistfeld, *The Doctrinal Unity of Alternative Liability and Market-Share Liability*, 155 U. PA. L. REV. 447, 449 (2007) (discussing the deterrent effect of the MSL doctrine).

67. See HOLMES, *supra* note 65, at 36–46.

68. *But see* Ronen Perry, *The Role of Retributive Justice in the Common Law of Torts: A Descriptive Theory*, 73 TENN. L. REV. 177, 188–92 (2006) (arguing that retributive justice has an influence on the development of tort law doctrines).

69. See FED. R. EVID. 403, 413, 414.

70. *Id.*

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,⁷¹ the prosecution can bring evidence of other crimes, wrongs, or acts that can be attributed to the defendant to establish motive, opportunity, intent, preparation, plan, knowledge, identity, or absence of mistake or accident.⁷² This evidence cannot be used to prove the defendant's bad character and courts are required to instruct the jury accordingly.⁷³ Interestingly, under Rule 404(b), as interpreted by the 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.⁷⁴ Judge Easterbrook has interpreted the Rule not as one of admissibility, because it "says that evidence 'may' be admissible for a given purpose, not that it is automatically admissible."⁷⁵

The similar-crimes doctrine, adopted in Rules 413 and 414 of the Federal Rules of Evidence, applies to sexual assault and

71. See *id.* 404(b). See also *United States v. Woods*, 484 F.2d 127, 137 (4th Cir. 1973), where the court stated prior to the enactment of Rule 404(b):

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.

72. FED. R. EVID. 404(b).

73. See *People v. Quinn*, 486 N.W.2d 139, 140 (Mich. Ct. App. 1992) ("Where, however, evidence of a defendant's other wrongful acts has been admitted for the limited purposes allowed under MRE 404(b), the prosecutor deprives the defendant of a fair trial in arguing that the jury should consider the evidence as substantive evidence of the defendant's guilt."); see also *Huddleston v. United States*, 485 U.S. 681, 689–92 (1988) (holding that the trial court is not required to make a preliminary finding that the petitioner 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, e.g., *Barnes v. City of Cincinnati*, 401 F.3d 729, 741–42 (6th Cir. 2005) (ruling that a statement made by a high-ranking official regarding lesbians in the city's police department was admissible under Rule 404(b)).

74. See, e.g., *Dowling v. United States*, 493 U.S. 342, 348–49 (1990) (holding that testimony tending to prove that the defendant had committed a crime, which had been brought in a prior trial that ended in acquittal, was rightly admitted under Rule 404(b) by the court in a subsequent trial because it established the defendant's identity).

75. *United States v. Jones*, 455 F.3d 800, 810 (7th Cir. 2006) (Easterbrook, J., concurring).

child molestation offenses.⁷⁶ 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.”⁷⁷

The superficial similarity between the pattern-of-behavior doctrines and the APP stems from their shared feature, namely that all three consider the past behavior of the defendant and affirm that past behavior influences the likelihood of conviction.⁷⁸ 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 or have actually committed the offense of which that person is presently accused. The defendant’s 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 simply based on the truism that the probability that a person committed one of two offenses (A or B) is greater than the probability that he committed A and greater than the probability that he committed B (unless there is full interdependence between the two offenses). The APP is not based on any interdependence between the offenses attributed to the defendant: the probability that he committed one offense does not change the probability that he committed another. Rather, only the probability that he committed *an unspecified offense* is affected.⁷⁹

76. FED. R. EVID. 413, 414.

77. *Id.* Under Rule 415 of the Federal Rules of Evidence this doctrine is also applicable to civil cases involving sexual assault and child molestation. 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 . . .”).

78. As Example 2 illustrates, sometimes the APP leads to acquittal rather than to conviction. See *supra* Example 2.

79. The Racketeer Influenced and Corrupt Organization Act (RICO), 18 U.S.C. §§ 1961–1968 (2006), can be interpreted as a tool for punishing individuals for unspecified offenses. Under RICO, a person who is a member of an enterprise that has committed any two specified crimes within a ten-year period can be charged with racketeering. *Id.* 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

2. Interdependence

Understanding the relationship between interdependence and the APP is crucial for setting the APP's scope and limits. As seen in this Section, interdependence may (although need not be) a reason not to apply the APP.

To illustrate the difference between the APP and the pattern-of-behavior doctrines and 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: pickpocketing and rape.⁸⁰ In this scenario, it is quite obvious that the pattern-of-behavior doctrines are not applicable, as there is no reason to believe that a person who committed pickpocketing would also commit rape. But, assume now that the two offenses are sexual assaults. In this case, 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 intended to do so.⁸¹ The evidence relating to each of the two charges would then bolster the case against the defendant with respect to the other charge.⁸² Indeed, a defendant who committed a sexual assault in the past is said to be more likely to have committed a later act of sexual assault.⁸³

criminal activity committed by those convicted under RICO their engagement in other activities—activities that have not been detected or 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 those of the APP. 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 conviction in the present case. Under the APP, in contrast, as illustrated by Example 2, the fact that a person was convicted of several offenses in the past decreases the probability of conviction in a later case. *See supra* Example 2. However, we do not suggest applying the APP across different trials. *See infra* Part IV.D.

80. *See supra* Example 1.

81. FED. R. EVID. 403, 413, 414.

82. *See, e.g.,* *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").

83. *See* Jodi Leibowitz, Note, *Criminal Statutes of Limitations: An Obstacle to the Prosecution and Punishment of Child Sexual Abuse*, 25 CARDOZO L. REV. 907, 939 n.127 (2003) ("For any sexual abuser, the likelihood that he has performed a similar abuse in the past—and that he will repeat it in the future—is extremely high.").

Assume, however, that even with the application of the prior-acts and similar-crimes doctrines, none of the charges can be proven beyond a reasonable doubt. Suppose that, for each of the charges in the modified version of Example 1, 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 while the threshold for conviction is .95. The APP is necessary to guarantee conviction in such a case. But, as is shown below, it is unclear whether the APP would be applicable.

One central consideration in determining whether the APP should apply 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 by supplementing the pattern-of-behavior doctrines or as an alternative to them.

Example 3, a 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 offense 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-a-reasonable-doubt standard is .95. Should the court use the APP and 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 also likely that he framed him 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 close to zero. Consequently, the

interdependence between the two offenses attributed to the defendant precludes conviction.

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

Example 4—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. Assume now that after applying the pattern-of-behavior doctrines, that probability increases to .9. The reason the court is not fully persuaded that the defendant committed each one of the offenses is because it is not clear that the complainants did not give their consent. The absence of victim consent is a precondition for convicting the defendant under prevailing law. At the same time the court is completely persuaded that the defendant committed all the acts attributed to him and also that he had the requisite intention for committing both sexual assaults. Assume that the required probability for satisfying the beyond-a-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 APP should be applied in this situation. Since the court's doubts with respect to each charge are independent of one another (the victim in each case is different and 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 assaults 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, even if the doubts 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 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 of the offenses is that, in each case, there is a different lone eyewitness whose reliability is questionable. Assume that the probability required for the beyond-a-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 apply. The reason for this is that, in this Example, using the APP becomes too complicated. If one eyewitness in the Example is a liar (or had made a mistake), then not only would acquittal of the relevant offense be justified, but also 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 APP would yield a probability of .91 that the defendant committed at least one of the alleged offenses.⁸⁴ This, by itself, would not be grounds for conviction. But, a different outcome would be obtained if, instead of .7, the probability of the defendant's guilt in each charge when examined separately was .8.⁸⁵

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 is true regardless of whether the pattern-of-behavior doctrines are applied.⁸⁶

84. $1 - (.3)^2 = .91$.

85. $1 - (.2)^2 = .96$.

86. See *supra* 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.⁸⁷

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.⁸⁸

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.⁸⁹

These conclusions support broad application of the APP. As long as the pattern-of-behavior doctrines are not applied and the doubt with respect to each charge has a different source, interdependence of the offenses does not preclude the application of the APP. Since the pattern-of-behavior doctrines have quite narrow application under prevailing law, the APP will seldom be precluded because of interdependence. 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—as demonstrated in Example 4—that meet the condition of a lack of interdependence between the different charges after application of the pattern-of-behavior doctrines. Other examples of this type are those in which the defendant's guilt in the offenses with which he is charged 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 at 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 intent of com-

87. See *supra* Example 5.

88. See *supra* Example 1 and modified Example 4 where the pattern-of-behavior doctrines are not applied.

89. See *supra* Example 4.

mitting murder. But, in each case, there is .1 probability that someone 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.⁹⁰

III. THE CASE FOR THE AGGREGATE PROBABILITIES PRINCIPLE

This Part provides reasons why the APP is a desirable rule and ought to be used by courts. There are three main advantages of the APP: First, the APP minimizes adjudication errors and makes the criminal law system more coherent, fair, and less discriminatory in treating those errors. Second, the APP is conducive to deterrence. Third, the APP reduces the costs of enforcement.

A. ADJUDICATION ERRORS

Adopting the APP in criminal law will likely increase the number of errors in convicting the innocent (false positives or Type I errors) and decrease the number of errors in acquitting

90. Jeremy 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. JEREMY BENTHAM, *Of the Proportion Between Punishments and Offences*, in AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION 165, 170 (J. H. Burns & H. L. A. Hart eds., 1970). 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* offence to which the punishment is to be annexed, but also of such *other* offences of the *same sort* as the offender is likely to have already committed without detection." *Id.* 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. *Id.*

the guilty (false negatives or Type II errors).⁹¹ 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.⁹² It is beyond the scope of this Article, however, to examine comprehensively the optimal mix of the two types of errors in criminal cases.⁹³ Let us assume that the beyond-a-reasonable-doubt standard implies a probability of at least .95 that the defendant committed the crime. Under this assumption, the law prefers setting free eighteen—but not nineteen—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. Adhering to the .95 threshold would mandate convicting the defendant for two offenses.⁹⁴ If the legal system acquits a person of all four offenses (as required by the DPP)⁹⁵ it will in effect endorse a principle under which it is better to have 9999 guilty people acquitted than a single 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 for people accused of committing a series of offenses over people accused of committing a single offense.⁹⁶ The requisite probability of guilt necessary to convict a defendant in the former case (the case involving a series of offenses) is much higher than what is required in the latter case (the case involving a single offense).⁹⁷ A person who has committed an offense beyond a reasonable doubt will be set free simply because the particular offense that he committed could not be proven beyond a reasonable doubt. The APP requires that the presumption of innocence should apply in a nondiscriminatory way.

91. See A. Mitchell Polinsky & Steven Shavell, *The Theory of Public Enforcement of Law*, in 1 HANDBOOK OF LAW AND ECONOMICS 403, 427–29 (A. Mitchell Polinsky & Steven Shavell eds., 2007) (discussing different ways to optimize Type I and Type II errors in law enforcement); STEIN, *supra* note 16, at 141–71 (discussing the allocation of risks of error in the law of evidence); I.P.L. Png, *Optimal Subsidies and Damages in the Presence of Judicial Error*, 6 INT'L REV. L. & ECON. 101, 102–04 (1986) (discussing different ways to optimize Type I and Type II errors in law enforcement).

92. See *infra* text accompanying note 99.

93. See *supra* note 91.

94. See *supra* note 4.

95. See *supra* Example 1.

96. See *id.*

97. See *id.*

As stated at the outset of this Part, the APP can be expected to raise the number of erroneous convictions of the innocent. But 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. In fact, even the *number* of errors resulting from convicting the innocent may be lower.

Suppose there is a constraint on the total amount of punishment the state can inflict on offenders, such as 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 prison for conviction on one charge. As a result, offenders who are convicted of several offenses would, on average, be sentenced to more years in prison, and other offenders—those who are charged with having committed one offense—would, on average, be sentenced to fewer years in prison. If the probability of error in a finding of guilt with respect to offenders who are convicted of several offenses is lower than with respect to defendants who are charged with having committed one offense (which is very likely),⁹⁸ 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 the APP could decrease the total costs of convicting the innocent.⁹⁹

98. C.Y. Cyrus Chu et al., *Punishing Repeat Offenders More Severely*, 20 INT'L REV. L. & ECON. 127, 136 (2000).

99. Cf. Henrik Lando, *The Size of the Sanction Should Depend on the Weight of the Evidence*, 1 REV. L. & ECON. 277, 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 and less cost to society); Talia Fisher, *Rethinking the Bipolar Structure of the Criminal Verdict* 19–29 (Oct. 13, 2009) (unpublished manuscript), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1488345 (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). 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 RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* 228 (7th ed. 2007) (increasing punishment for repeat offenders is justified because the risk of convicting the innocent is lower in their case); Chu et al., *supra* note 98, at 135 (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).

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 time period that an offender is sent to prison for one offense at its current level under the DPP. This strategy would arguably decrease the number of errors in convicting the innocent, since the required probability of guilt for conviction would be higher under the APP than the DPP.

B. DETERRENCE

The APP is superior to the DPP on deterrence grounds, particularly for repeat offenders. Under the APP, repeat offenders have a smaller chance of avoiding conviction than under the DPP.¹⁰⁰ This implies a higher expected sanction for repeat offenders under the APP and, accordingly, greater deterrence than under the DPP.¹⁰¹

This advantage of the APP is significant if one assumes that the expected sanction necessary to achieve optimal deterrence is in fact higher for repeat offenders than for other offenders.¹⁰² But the APP is even more conducive to deterrence

100. See *supra* Part II.B.

101. See *infra* Part IV.B.

102. 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 David A. Dana, *Rethinking the Puzzle of Escalating Penalties for Repeat Offenders*, 110 YALE L.J. 733, 737 (2001) (arguing that declining penalties for repeat offenders are optimal since the probability of detection escalates with offense history); Winand Emons, *A Note on the Optimal Punishment for Repeat Offenders*, 23 INT'L REV. L. & ECON. 253, 254 (2003) (arguing that when punishment is a fine, under certain conditions, the optimal sanction scheme decreases); Winand Emons, *Escalating Penalties for Repeat Offenders*, 27 INT'L REV. L. & ECON. 170, 171 (2007) (arguing that under certain conditions, increasing sanctions for repeat offenders is optimal and, under other conditions, the reverse holds true); Thomas J. Miceli & Catherine Bucci, *A Simple Theory of Increasing Penalties for Repeat Offenders*, 1 REV. L. & ECON. 71, 72 (2005) (claiming that repeat offenders should be punished more severely than other offenders, because of their diminished employment opportunities); A. Mitchell Polinsky & Daniel L. Rubinfeld, *A Model of Optimal Fines for Repeat Offenders*, 46 J. PUB. ECON. 291, 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 case, less severely in another type of case, and with the same severity in other types of cases); A. Mitchell Polinsky & Steven Shavell, *On Offense History and the Theory of Deterrence*, 18 INT'L REV. L. & ECON. 305, 306–07 (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); Richard A. Posner, *An Economic Theory of the Criminal Law*, 85 COLUM. L. REV. 1193, 1215

given that repeat offenders are often “professionals,” whereas one-time offenders are often amateurs.¹⁰³ As professionals, repeat offenders are likely to be more sophisticated than their first-time counterparts. Indeed, professional criminals are more responsive to sanctions and also more inclined to take precautions to reduce the likelihood of conviction.¹⁰⁴ 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 brought to trial or ever convicted. They are extremely proficient at playing by the rules of the game and the DPP facilitates this effort. It is this class of criminals that will be particularly deterred under the APP.¹⁰⁵

Deterrence considerations would not favor the use of the APP across different trials in cases characterized by Example 2. In that Example, a defendant is charged with two offenses and the probability that he committed each offense is .95. In

(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”); Ariel Rubinstein, *On an Anomaly of the Deterrent Effect of Punishment*, 6 ECON. LETTERS 89, 90 (1980) (arguing that punishing repeat offenders more harshly increases deterrence of offenders).

103. See, for example, the “dangerous special offender” statute, 18 U.S.C. § 3575 (1988) (repealed 1984), which provided for an enhanced penalty of up to twenty-five years imprisonment for repeat offenders, professional criminals, and organized crime offenders.

104. See Lucian Arye Bebchuk & Louis Kaplow, *Optimal Sanctions and Differences in Individuals’ Likelihood of Avoiding Detection*, 13 INT’L REV. L. & ECON. 217, 223 (1993) (discussing optimal enforcement when some individuals are more sophisticated than others).

105. A possible counterargument 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* 11–16, 18–24 (Oct. 20, 2005) (unpublished manuscript), 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); cf. Chris William Sanchirico, *Character Evidence and the Object of Trial*, 101 COLUM. L. REV. 1227, 1276 (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).

such a case, we suggested that under the APP the person ought to be convicted in one rather than two offenses as the probability he committed both offenses is lower than .95. Yet, if applied to cases involving different trials, the APP 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. For this and other reasons¹⁰⁶ 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 typically increase.¹⁰⁷ 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 X. It is typically much harder—and more costly—to collect the tenth item of evidence than the ninth item, the eighth item, and so on.¹⁰⁸ 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 another offense Y, reasonably attributed to the defendant.

IV. OBJECTIONS

This Part examines several objections to the APP. These objections should be taken seriously. Yet, none of them provide

106. See *infra* Part IV.D.

107. See *infra* note 108.

108. See, e.g., 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 a reasonable doubt’”); 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.”); Ehud Guttel & Alon Harel, *Matching Probabilities: The Behavioral Law and Economics of Repeated Behavior*, 72 U. CHI. L. REV. 1197, 1213 n.54 (2005) (stating that increasing enforcement can be achieved by either requiring enforcers to work more or recruiting additional personnel, under either approach “the marginal cost of enforcement is likely to increase”); 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.”).

a sufficient reason to reject the APP. Instead, awareness of the force of these objections results in certain modifications of our proposal and in narrowing the scope of the application of the APP.

This Part ignores, however, one doctrinal 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.”¹⁰⁹ Under Rule 14(a) of the Federal Rules of Criminal Procedure, courts may order separate trials if the joinder of offenses appears to prejudice a defendant or the government.¹¹⁰ 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 require alteration. Applying the APP would require that different offenses which, under Rule 14(a) are currently investigated in different trials, be joined together.

A. MANIPULATIONS BY THE PROSECUTION AND AGENCY COSTS

Arguably, the use of the APP will invite large-scale abuse.¹¹¹ After all, it is relatively easy to bring *some* evidence

109. FED. R. CRIM. P. 8(a).

110. *Id.* 14(a).

111. 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) (Brennan, J., dissenting) (“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.”). The risk of prosecutorial misconduct is a consideration in shaping procedural and evidentiary doctrines. See Charles P. Bubany & Frank F. Skilern, *Taming the 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); Angela J. Davis, *Prosecution and Race: The Power and Privilege of Discretion*, 67 FORDHAM L. REV. 13, 20–25 (1998) (discussing prosecutors’ vast discretion and power); Angela J. Davis, *The American Prosecutor: Independence, Power, and the Threat of Tyranny*, 86 IOWA L. REV. 393, 410–15 (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); Bruce A. Green, *Policing Federal Prosecutors: Do Too Many Regulators Produce Too Little Enforcement?*, 8 ST. THOMAS L. REV. 69, 70 (1995) (“[E]vidence 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).

indicating the guilt of any defendant. Consequently, there is a grave risk that under the APP, any person could be convicted for some offense without significant evidence supporting the conviction. The prosecution would find it 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 a reasonable doubt, the prosecutor could easily collect *some* evidence suggesting that the defendant committed a different offense and thus overcome the evidential hurdles to conviction. It seems therefore that there are serious institutional reasons to stick with the current procedural system governed by the DPP.

This objection is not a reason for rejecting the APP, but rather for designing it in a way that would alleviate the manipulation concerns. Recall that under the principles of evidence law, a person cannot be convicted on the basis of statistical evidence alone.¹¹² The APP does not change the rules precluding conviction for statistical evidence.¹¹³ To convict a person, case-specific evidence must be brought with respect to each of the relevant charges.¹¹⁴ Otherwise, conviction is not possible.

This requirement may not fully alleviate the concern. In particular, it does not preclude the possibility of collecting low probability evidence for conviction. To prevent abuse of the system, 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-the-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 admitting evidence is low.

A related objection is based on the premise that it is better to convict a person of a specified offense rather than an unspecified one. Yet, under the APP, prosecutors may fail to act accordingly and they may prefer conviction for an unspecified offense.

This objection is unpersuasive. At least in terms of efficiency-based considerations, there is no clear reason to prefer conviction for a specified offense over a conviction for an unspecified one. Second, *even if* it is better to convict a defendant for a

112. See *infra* Part IV.B.

113. See *infra* Part IV.B.

114. STEIN, *supra* note 16, at 154.

specified offense, there is no reason to assume that prosecutors—who are the public’s fiduciaries—will ignore this fact and will not take it into account when deciding whether to use the APP in a particular case. Third, if prosecutors are expected to ignore this consideration, and consequently over-use the APP, this misuse of prosecutorial discretion could be addressed directly by the prosecutors’ superiors who could issue clear instructions as to the APP’s appropriate use. Other institutional solutions to overcome this abuse could be developed as well.

Arguably prosecutors may bring several charges of low probability against defendants not only to economize enforcement costs, but also to create a heavy and presumably unjust burden for the defendant to rebut the charges. Again, this concern should be addressed directly by the courts, as is often done in cases of abuse.¹¹⁵ For example, courts could reduce sanctions imposed upon defendants when the prosecution brings charges that fail to satisfy the preponderance-of-the-evidence standard. That would disincentivize the prosecution from bringing charges of low probabilities in order to impose unjust burdens on the defendant.

B. STATISTICAL EVIDENCE AND STATISTICAL INFERENCES

Arguably, the APP is based on “statistical evidence,” evidence that is often regarded as inadmissible.¹¹⁶ Moreover, establishing the defendant’s guilt under the APP is based on a probabilistic conception of the beyond-a-reasonable-doubt standard, which, according to some scholars, undermines the trust of the public in the criminal justice system.¹¹⁷

In discussing the difference between “naked statistical evidence” and “trace-based forms of evidence,” a leading theorist of evidence law 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

115. Cf. Andrew James McFarland, Note, *Lewis v. United States: A Requiem for Aggregation*, 46 CATH. U. L. REV. 1057, 1077–79 (1997) (discussing cases in which defendants subjected to multiple charges for petty offenses were held to be entitled to the additional protections offered by jury trial).

116. See *infra* note 121 and accompanying text.

117. See Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts*, 98 HARV. L. REV. 1357, 1366–67 (1985).

was struck by a 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.¹¹⁸

Statistical evidence does not focus on the individual defendant; instead, it purports to establish guilt by using generalizations concerning the group to which the individual defendant belongs, e.g., his race, religion, or any other indicator correlated with criminal activity.¹¹⁹ Thus, if statistical evidence is used, the fact that crimes of a particular type are committed by individuals of a certain background could be used to support conviction. 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.”¹²⁰ An individual defendant who is charged with fraud or 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. This evidence is based exclusively on statistical inferences rather than on case-specific information and therefore there is nothing that the individual defendant can do to defeat it.¹²¹ Even under this framework, one can show that an opposition to the use of naked statistical evidence 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.¹²² Therefore, it is arguably the case that statistical inference, rather than case-specific evidence led to the conviction. There is no connection between the inference leading to conviction and the particular

118. STEIN, *supra* note 16, at 206–07 (citation and footnotes omitted).

119. See, e.g., Eric S. Janus & Robert A. Prentky, *Forensic Use of Actuarial Risk Assessment with Sex Offenders: Accuracy, Admissibility and Accountability*, 40 AM. CRIM. L. REV. 1443, 1448–58 (2003) (discussing the use of statistical evidence in evaluating risks related to sexually violent predators).

120. STEIN, *supra* note 16, at 206–07.

121. *Id.* at 207. It is beyond the scope of this Article to discuss the pros and cons of using naked statistical evidence in criminal cases. Cf. Henry M. Hart, Jr. & John T. McNaughton, *Evidence and Inference in the Law*, in EVIDENCE & INFERENCE 48, 54 (Daniel Lerner ed., 1958) (“[T]he law refuses to honor its own formula when the evidence is coldly ‘statistical.’”); Nesson, *supra* note 117, at 1379 (stating that cases based only on probabilistic evidence are unlikely to reach the jury because “the factfinder cannot reach a conclusion that the public will accept as a statement about what happened”).

122. See *supra* Part I.

circumstances of the defendant, since any defendant facing similar charges would be convicted. Yet, unlike the paradigmatic case of statistical evidence presented above, 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 each offense—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. In both cases there are several separate items of evidence supporting the charge, each of which is not sufficient in itself for conviction, but can, in the aggregate, sustain a conviction.

Indeed, a conviction under the APP is typically based on the accumulation of all case-specific evidence: all the evidence brought against the defendant with respect to each charge must be sufficient to establish beyond a reasonable doubt that the defendant committed at least one of the crimes. A comparison of two cases demonstrates this point. In one case, a certain amount of evidence is required for conviction under the DPP, for example, ten pieces of evidence. In the second case, a certain amount of evidence is required for conviction under the APP, for example, eight pieces of evidence for one of the offenses and eight pieces of evidence for another 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 is that the probabilistic conception of the beyond-a-reasonable-doubt standard, upon which the APP rests, weakens public trust in the criminal justice system.¹²³ According to the most prominent advocate of the public trust argument, “[o]ur criminal justice system seeks to produce authoritative finality by inducing the general public to defer to jury verdicts.”¹²⁴ To achieve this goal,

[T]he 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 . . . [S]o long as the evidence prevents specific quantification of the degree of that uncertainty, an out-

123. See Nesson, *supra* note 117, at 1366–67.

124. Charles Nesson, *Reasonable Doubt and Permissive Inferences: The Value of Complexity*, 92 HARV. L. REV. 1187, 1195 (1979).

side observer has no reasonable choice but to defer to the jury's verdict.¹²⁵

The use of the APP does not conflict with these concerns. There are no grounds for concern that applying the APP would undermine public trust in the criminal justice system. This is because the APP would enable jurors and judges to conclude that it has been proven beyond a reasonable doubt that one of the offenses was committed by a particular defendant, even if it cannot be established which specific offense was committed. Explicit probability calculus—which according to this argument weakens the public trust in the courts—is no more necessary in applying the APP than in the DPP when several pieces of evidence must be 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 courts to consider even relatively low probability offenses when determining guilt or innocence. APP opponents may maintain that its use would encourage the prosecution to bring as much evidence as it can reasonably amass with respect to any seemingly criminal behavior. It follows 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, this is a trivial concern. First, as discussed above, adopting a minimal threshold that precludes courts from aggregating low probability offenses 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 increase the threshold for conviction (thereby reducing the amount of litigation) than to reject the APP altogether.¹²⁶ Third, although the APP would stimulate more litigation, it is likely that the litigation costs would decrease. 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.¹²⁷ 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

125. *Id.* at 1199 (citation omitted).

126. *Cf. supra* Part III.A.

127. *See supra* Part III.C.

to be higher on average than the cost of increasing this probability from, say, .5 to .55 (as allowed under the APP). Lastly, the litigation generated by the APP is not frivolous. On the contrary, as was shown earlier, this increase in litigation would result in a correlative increase in justified convictions and better enforcement of the law.¹²⁸

D. IMPLEMENTATION DIFFICULTIES AND THE MEANING OF A PROBABILISTIC THRESHOLD

The APP might be difficult to apply in practice as courts could make mistakes. A major difficulty of the APP is the interdependence of the offenses with which the defendant is charged.¹²⁹ Given the risks of interdependence, the application of the APP may become too hard for both the courts and jurors. Furthermore, sometimes the interdependence is hidden.¹³⁰ 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 jurors.

The interdependence problem was discussed above and we established that so long as the doubts differ with respect to each offense and the pattern-of-behavior doctrines are not applied, there would be no reason not to apply the APP, regardless of the lack or presence of interdependence.¹³¹ This would encompass a large range of cases 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 the separate 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 (as in Example 4),¹³² there is no reason not to apply the APP. Judges would have no difficulty deciding whether these conditions have been met in any given case and, accordingly, could instruct the jury on whether to apply the APP.

The problem of hidden interdependence warrants special attention. Hidden interdependence precludes application of the APP. For example, suppose a driver is caught five times for

128. See *supra* Part III.B.

129. See *supra* Part II.B.2.

130. See *infra* text accompanying notes 132–33.

131. See *supra* Part II.B.2.

132. See *supra* Example 4.

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 (because it is calibrated every night) and that the rate of error is especially high with brightly colored cars. Imagine that our driver was caught on five occasions in the evening and his car is brightly colored. Assuming the court is unaware of the radar's defects, aggregating the probabilities could create a 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 applying the APP. Furthermore, given awareness of the problem of hidden interdependence, courts would be more likely to recognize the need to diligently obtain information concerning cases involving probabilistic interdependence.

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.¹³³ To be sure, under the DPP, courts should be able to judge whether the beyond-a-reasonable-doubt standard has been satisfied. A court's determination of whether the standard has been satisfied has some probabilistic features.¹³⁴ But under the DPP, courts are not required to ascribe accurate probabilities to their findings.¹³⁵

This objection does not justify rejecting the use of the APP across different offenses in the same trial. Admittedly a court using the APP should look also at the complete picture, namely, all charges against the defendant. But the APP does not require courts to ascribe precise probabilities to each offense. Rough terms such as high probability would suffice to justify the use of the doctrine. The question that the judge or jury would have to answer is whether given the evidence there is a very high likelihood that the defendant committed an offense (rather than a particular offense as required under the DPP). If the judge (or jury) concludes that such evidence exists, then a conviction would follow.

133. See STEIN, *supra* note 16, at 64 ("In real life, evidence is constantly missing . . . Fact-finders have to settle for less.")

134. See *supra* text accompanying notes 16–17.

135. See STEIN, *supra* note 16, at 65 (noting that fact-finders are forced to make probability estimates based on inadequate evidence).

To understand just what the APP requires of courts, it is useful to compare its requirements to those under the DPP. The DPP requires 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.¹³⁶ Only if there is sufficient evidence to establish guilt in one of these offenses is the defendant convicted of that offense.¹³⁷ In contrast, under the APP, the court must address the additional question of whether there is sufficient evidence that the defendant committed any one of the three charged offenses.¹³⁸ Thus, using the APP, a court could conclude that even though it cannot convict the defendant for committing any specific offense, it can convict him for committing one indeterminate offense because there is no reasonable doubt that he committed one offense. This shows that there is no meaningful difference between how the reasonable doubt principle is applied under the APP and under the DPP. While under the DPP, the court convicts the defendant when there is no reasonable doubt he committed the specific offense attributed to him, under the APP, the court convicts the defendant 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 used by courts implicitly,¹³⁹ and thus there is no need to recognize it

136. See *supra* note 1 and accompanying text.

137. *Id.*

138. 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. See COOTER & ULEN, *supra* note 32, at 351–52 (“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.”). 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 *id.*; 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).

139. See, e.g., Andrew D. Leipold & Hossein A. Abbasi, *The Impact of Joinder and Severance on Federal Criminal Cases: An Empirical Study*, 59 VAND. L. REV. 349, 367 (2006) (showing empirically that criminal defendants

explicitly. Moreover, if courts are applying it implicitly, forcing them to apply it explicitly may result in double counting. This objection to the use of the APP contends that when several charges are brought against a defendant—even if unrelated to one another—the judge and the jury are in fact influenced by the accumulation of charges and tend to convict more readily than if there were only one charge.¹⁴⁰

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 the APP is a desirable mechanism that provides a justification for trying several unrelated charges, even of different natures, against one defendant in the same trial.¹⁴¹ Of course, there may be other considerations concerning joining different charges in one trial, which could have great weight, but the desirability of the APP should also be regarded as a relevant factor in making the procedural decision whether to charge a defendant for different crimes in one trial or not.¹⁴²

Is it true that courts implicitly aggregate probabilities across offenses? It is hard to know whether this is empirically correct. With respect to judges, it may be possible to assume a certain commitment on their part to examine each charge separately. The prevailing legal ethos founded on the DPP principle does not allow considerations of the type examined above.¹⁴³ 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. Further-

who face multiple charges in a single trial have a harder time prevailing than those who face several trials of one count each).

140. *Id.*

141. The prior-acts and similar-crimes doctrines allow courts and jurors, under certain conditions, to consider the accumulation of the evidence of all charges. *See supra* Part II.B.1. But as we explained previously, these two doctrines differ from the APP. *Id.*

142. *See supra* Part IV. 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.

143. *See* Nesson, *supra* note 124, at 1188 (“[D]ue process requires that the prosecution in a criminal case prove each and every material element of a criminal offense beyond a reasonable doubt . . .”).

more, 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 inconsistent intuition of judges and jurors.

Arguably, even if judges do not apply the APP, it is possible that the police and prosecution apply some version of it in making their decisions regarding law enforcement efforts. According to this theory, 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. As such, they generally have more information about his potential involvement in perpetrating crimes and they try harder to collect even more evidence to increase the chances of conviction.¹⁴⁴

This argument is unpersuasive. First, if courts refuse to apply the APP it will be a factor in prosecutors' decisions not to charge suspects even if, when aggregating the probabilities, they are convinced that the defendant 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 it is still not clear why courts should not apply the APP. By refusing to adopt the APP, courts encourage the prosecution to incur excessive enforcement costs as such costs under the APP are lower than under the DPP.¹⁴⁵

Finally, it could be argued that, at least in plea bargains, the APP is already applied in practice. When there are several accusations against a defendant (even when none of them meets the threshold necessary for conviction) the mere cumulative force of the different allegations influences the nature of the deal made between the prosecution and defendant. Even assuming this argument is correct, courts should apply the APP. As described above, prosecutors act in the shadow of the prospective trial.¹⁴⁶ Therefore, if the APP is not applied by

144. Dana, *supra* note 102, at 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 face huge obstacles in developing any comparative assessments of the success of different groups of offenders in evading detection.”).

145. See *supra* Part III.C.

146. See Oren Gazal-Ayal, *Partial Ban on Plea Bargains*, 27 CARDOZO L. REV. 2295, 2325 (2006) (noting that prosecutors incorporate rules which ex-

courts, this will most certainly affect the shape of plea bargains. Defendants would not accept deals when the trial (governed now by the DPP) is unlikely to result in conviction. Furthermore, even if the APP is perfectly applied in the context of plea bargains, there is still no reason why the APP 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 argued that deterrence-based theories, particularly theories that focus on efficiency, would likely endorse a moderate version of the APP. It is time to consider the justifiability of the APP and DPP from the perspective of justice-based theories. 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.

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*.”¹⁴⁷ This observation lies at the foundation of many retributivist theories.¹⁴⁸ David Dolinko conceives of retributivists as those who justify punishment “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.”¹⁴⁹ Under what Dolinko labels “bold retributivis[m],” “lawbreakers deserve punishment and that this, all by itself, constitutes a good or sufficient reason for the state to inflict punishment on

clude reliable evidence in the plea bargaining stage because the bargaining “takes place in the shadow of the trial”).

147. IMMANUEL KANT, *THE METAPHYSICS OF MORALS* 105 (Mary Gregor ed., Mary Gregor trans., Cambridge University Press 1996) (1797).

148. We do not argue, however, that Kant was committed to the versions of retributivism that we present below. Kant’s theory of punishment has been interpreted by many theorists, and we do not purport to provide an interpretation of it here.

149. David Dolinko, *Some Thoughts About Retributivism*, 101 *ETHICS* 537, 541–42 (1991).

them.”¹⁵⁰ Dolinko also asserts that retributivists believe in proportionality, namely, that wrongdoers ought to be made to suffer in proportion to their offenses.¹⁵¹

Criminals, according to this view, simply deserve to be punished, and this desert provides the justification for inflicting punishment on them.¹⁵² Hence, under this version of retributivism, if it is shown that an agent committed a wrong there is a reason to impose a sanction on that person, even if the nature of the wrong remains unspecified. It should be noted that there is no consensus as to what retributivism really is,¹⁵³ but for the purposes of the discussion in this Part, we focus on Dolinko’s version. Different characterizations of retributivism are unlikely to lead to a different result.

In contrast, expressivists underscore the importance of the expressive, educational, and communicative aspects of criminal sanctions.¹⁵⁴ Under expressivist theories, sanctioning a wrongdoer is a public manifestation of condemnation and disapprobation of his deeds.¹⁵⁵ Some believe that the need for condemnation is in itself a 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.¹⁵⁶

Robert Nozick falls into the former camp. He argues that “[r]etributive punishment is an act of communicative behavior.”¹⁵⁷ In elaborating on the concept of communicative behavior, Nozick speaks of retributive principles as encompassing two goals. The first is to “connect the wrongdoer to value qua value,” and the second is to connect the wrongdoer in a way

150. *Id.* at 542.

151. *Id.* at 550 (discussing Jean Hampton’s characterization of retributivism in her essay *The Retributive Idea*, in FORGIVENESS AND MERCY 125–26 (1988)); cf. Thomas E. Hill, *Kant on Wrongdoing, Desert, and Punishment*, 18 LAW & PHIL. 407, 409 (1999) (illustrating retributivists’ use of proportionality in sentencing).

152. Hill, *supra* note 151, at 425.

153. See Russell L. Christopher, *Detering Retributivism: The Injustice of “Just” Punishment*, 96 NW. U. L. REV. 843, 845 n.1 (2002) (“[A] precise definition of retributivism has proven elusive . . .”).

154. See *infra* notes 155–60 and accompanying text.

155. See, e.g., Joel Feinberg, *The Expressive Function of Punishment*, in DOING AND DESERVING 95, 98 (1970); Jean Hampton, *The Moral Education Theory of Punishment*, 13 PHIL. & PUB. AFF. 208, 212 (1984).

156. See Dolinko, *supra* note 149, at 541–42 (comparing justifications for punishment).

157. ROBERT NOZICK, PHILOSOPHICAL EXPLANATIONS 370 (1981).

“that value qua value has a significant effect in [the criminal’s] life, as significant as his own flouting of correct values.”¹⁵⁸ Other eminent expressivist theorists of punishment defend related justifications. Joel Feinberg asserts that “punishment is a conventional device for the expression of attitudes of 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.”¹⁵⁹ 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.”¹⁶⁰ All of these theorists justify punishment in expressivist terms—as a means of conveying and expressing condemnation.

Although expressivist theories would not 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*. Arguably, it is a prerequisite for conveying condemnation and disapproval of an act to identify clearly the object of condemnation and disapproval, i.e., to identify unambiguously the condemned act.¹⁶¹ Punishing a person for an offense that person may or may not have committed rather than for the offense that person actually committed dilutes the important expressive, educational, and communicative messages of punishment. Hence, expressivist theories would likely reject the APP because, under it, 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 not have committed simply because the act is part of a disjunction of acts diverges significantly from condemning a specific act. Only condemning a person for a specific act can meet expressivist concerns.

The rejection of the APP by expressivist theorists is no accident. It is a by-product of the way these theorists address what seems to be one of their apparent weaknesses. Expressiv-

158. *Id.* at 376–77.

159. Feinberg, *supra* note 155, at 98.

160. Hampton, *supra* note 155, at 212.

161. 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. *See, e.g., id.* at 225 (discussing that punishment should educate someone that a particular act is wrong and not concern itself with their character or moral duties).

ist theories are vulnerable to the accusation that condemning theft, rape, or murder does not necessitate the infliction of sanctions on the perpetrator.¹⁶² 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.¹⁶³ The distinct nature of punishment as an expressive practice requires that the object of condemnation be specific and concrete.¹⁶⁴ Punishment must, therefore, be designed to express disapproval of a particular act that was performed by the perpetrator, not for an act that may or may not have been committed by him. Evidence of this requirement for specificity is abundant in expressivist theories.¹⁶⁵ 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.”¹⁶⁶ Feinberg also maintains that “punishment surely expresses the community’s strong *disapproval* of what the criminal did.”¹⁶⁷ Communicating disapproval by punishing an individual for a disjunction of acts, e.g., stealing or committing fraud, does not satisfy the specificity of expressive condemnation required by these theories.

Proponents of the APP could counter the expressivist arguments by advocating 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. 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 does not capture the significance of the condemnation of a concrete act—the pre-

162. See, e.g., *id.* (acknowledging that punishment is only one form of moral education).

163. See Feinberg, *supra* note 155, at 98–99, 263; Hampton, *supra* note 155, at 225.

164. Hampton, *supra* note 155, at 216 (“[O]ur principal concern as we punish is to get the wrongdoer to stop doing the immoral action by communicating to her that her offense was immoral.”).

165. See, e.g., *id.* at 225 (discussing how punishment should focus on educating that a specific act is wrong).

166. *Id.* at 227 (emphasis added).

167. Feinberg, *supra* note 155, at 100.

cise 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.

Another reason why expressivist concerns may lead to rejection of the APP is the well-being of crime victims. 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 may not be convicted of any specific crime and, consequently, no victim of a particular crime can conclusively establish that a wrong has been committed against him.

In summary, 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 APP, 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.

B. THE CASE FOR A MODERATE AGGREGATE PROBABILITIES PRINCIPLE

When, if ever, should courts recognize and employ the APP? For retributivists (at least those advocating the type of retributivism described above), the answer would be that the APP should be applied without limit. For expressivists, it depends on whether using the APP will effectively serve the expressive, educational, and communicative functions of criminal law. It seems 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 more these theorists would wish to apply the DPP. 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 a 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. Some criminal law theorists acknowledge this concern as “fair labeling.”¹⁶⁸ 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.”¹⁶⁹

But even in cases in which a person is charged with two offenses of the same type, e.g., two instances of fraud, it seems implausible under expressivist theories to convict that person of an unspecified offense. 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 the defendant since the disjunction of the offenses consists of very different acts.

Yet, there are circumstances under which the heterogeneity of the situations should not bar an unspecified conviction under 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 a 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 specific enough and to convey a clear and concrete disapproval of the behavior in question.

168. See, e.g., A. P. Simester & G. R. Sullivan, *On the Nature and Rationale of Property Offences*, in *DEFINING CRIMES* 168, 186–87 (R. A. Duff & Stuart P. Green eds., 2005).

169. *Id.*

Another parameter that seems to bear on 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 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 the defendant committed.

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.¹⁷⁰ Yet, even the latter need not reject the APP outright. 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 nature of criminal law. The reluctance of the legal system to endorse the APP suggests that expressivist concerns play an important role in this field.

CONCLUSION

This Article investigated a puzzle: why has the APP been unequivocally and universally rejected in criminal law? It is our claim that the reason is probably rooted in expressivist theories of punishment. These theories affirm that the act for which a person is condemned 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, practicalities would remain that would limit the scope of the APP's application. An appropriately modified version of the APP, which takes into account the objections discussed in this Article, would promote deterrence, minimize adjudication errors, and save enforcement costs. Given such modifications, the APP could be insulated from abuse and tailored to be consistent with justice-based theories of punishment, including expressivist theories.

We therefore suggest that the APP be applied, albeit with great caution and awareness of the difficulties it can generate. First, the APP should be applied only to charges brought in the same trial. Implementing the APP across trials could be diffi-

170. See *supra* Part V.A.

cult 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.¹⁷¹ 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 of the APP, especially in contexts where the risk of repeat offenders escaping punishment under the DPP is high.¹⁷² Third, in applying the APP, courts should pay particular attention to the interdependence problem.¹⁷³ Such sensitivity would be especially imperative when the pattern-of-behavior doctrines are applied by the court. Fourth, the APP should be applied only in those cases where the probability of the defendant's guilt is higher than .5.¹⁷⁴ This restriction would reduce the risk of abuse by the prosecution and at the same time would provide courts with a familiar standard of proof in which they are well-trained.

This Article is ultimately the by-product 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 commonsense 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. A rejection of the APP is one such rare case.

171. *See supra* Part IV.

172. *See supra* Part III.B.

173. *See supra* Part II.B.

174. *See supra* Part IV.A.