"LET 'EM PLAY"
A STUDY IN SPORTS AND LAW

Mitchell N. Berman*

INTRODUCTION: ONE STEP OVER THE LINE

Kim Clijsters was the feel-good story of the 2009 U.S. Open. A former world #1, the 26-year-old Belgian had been retired for 2 years, during which time she had married and borne a child, when she surprised the tennis world by announcing her return in the summer of 2009. Entering the tournament as an unranked wildcard, Clijsters defeated Denmark's Caroline Wozniacki in straight sets to become the first unseeded player ever to win the U.S. Open and the first mother to win a grand slam event in 30 years.

The match that made the headlines, however, was not the final. It was Clijsters's semifinal contest against the #2 seed, Serena Williams. Straight off her victory at Wimbledon, Williams had powered through the women's draw at the Open without losing a set; with the third-round defeat of #1 seed Dinara Safina, she was the odds-on favorite to win her third grand slam tournament of the year. Instead, Williams lost the first set 4-6 and found herself serving to Clijsters at 5-6 in the second. Down 15-30, Williams's first serve to the ad court was wide. On her second serve, the line judge called Williams for a foot fault, putting her down double-match point. At this, Williams exploded, walking over to the judge several times, gesticulating with her racket in an ominous manner, shouting, and threatening to do things with the ball that the lineswoman was bound to find unwelcome. Because Williams had already

* Richard Dale Endowed Chair in Law, Professor of Philosophy (by courtesy), The University of Texas at Austin.
Earlier drafts of this essay were presented at the 2010 Analytical Legal Philosophy Conference at NYU, and at faculty workshops at the University of Texas School of Law and the McMaster University Philosophy Department. I am grateful to participants at these events for their reactions and criticisms, to Ronen Avraham, David Enoch, Matt Kramer, Stefan Sciaraffa, Seana Shiffrin, Matt Spitzer, Abe Wickelgren, and Ben Zipursky for very helpful written comments, and to Rich Friedman for the title. Shane Anderson and Anthony Arguijo provided excellent research assistance.

committed a code violation earlier in the match for racket abuse, this second code violation called forth a mandatory one-point penalty. That single point gave the match to Clijsters.

Williams has few defenders. Her outburst was further penalized with a $10,000 fine that, after many commentators bemoaned its inadequacy, was raised to $82,500 and supplemented with a two-year probation. But without condoning or excusing Williams's response to the foot fault call, I'm interested in a different question: whether the call should have been made at all. CBS color commentator and former tennis great John McEnroe thought not. As he remarked at the time: "you can't call that there." His point was not that the call was factually mistaken,[1] but rather that, even assuming arguendo that it was factually supportable, it was an inappropriate call to make at that point in the match: the lineswoman should have cut Williams a little slack. Many observers agreed. As another former tour professional put it, a foot fault "is something you just don't call—not at that juncture of the match."[2]

The McEnrovian position—that at least some rules of some sports should be enforced less strictly toward the end of close matches—is an endorsement of what might be termed "temporal variance." It is highly controversial. Indeed, some people find it simply incredible that a call conceded to be correct at one time could be thought improper at another. As one letter writer to the *New York Times* objected: "To suggest that an official not call a penalty just because it happens during a critical point in a contest would be considered absurd in any sport. Tennis should be no exception."[3] On this view, which possibly resonates with a common understanding of what it means to follow "the rule of law," rules of sports should be enforced with resolute temporal invariance.

Now, perhaps McEnroe was wrong about the Williams foot fault. But the premise of the letter propounding the competing view—that participants and fans of any other sport would reject temporal variance decisively—is demonstrably false. Indeed, one letter appearing in *Sports Illustrated* objected to the disparity of attention focused on Williams as compared to U.S.

---

[1] The rule governing foot faults provides that, "During the service motion, the server shall not . . . [t]ouch the baseline or the court with either foot." International Tennis Federation, Rule of Tennis 18.c (2010). *See also* USTA Comment 18.3 (2007) ("A player commits a foot fault if after the player's feet are at rest but before the player strikes the ball, either foot touches . . . the court, including the baseline.").

[2] Michael Wilbon, *A Call and a Response That Can't Be Defended*, WASH. POST, Sept. 14, 2009, at D03, *available at* http://www.washingtonpost.com/wp-dyn/content/article/2009/09/13/AR2009091302533.html.

[3] Vince Bray, Letter to the Editor, N.Y. TIMES, Sept. 20, 2009 (Sports Sunday), at 5.

Open officials, precisely on the grounds that "[r]eferees for the NFL, NHL and NBA have generally agreed that in the final moments, games should be won or lost by the players and not the officials."[4] I am unsure just how general this supposed agreement is. But I'd warrant that most fans of professional basketball would affirm that contact that would constitute a foul through most of the game is frequently not called during the critical last few possessions of a close contest. Moreover, most fans I have spoken with believe this is not only how it is, but how it ought to be. In any event, precious few of those who disagree would contend that the status quo is absurd. So an insistence on rigid temporal invariance requires argument not just assertion.

However, advocates of temporal variance ought not to be too smug either. For while the negative import of temporal variance is clear—namely, the *denial* of categorical temporal *in*variance—its positive import is anything but. Surely those who believe that the foot fault ought not to have been called against Serena Williams in her match against Kim Clijsters mean implicitly to invoke a principle broader than "don't call foot faults in the twelfth game of the second set of semi-final matches in grand slam tournaments." But how much broader? Is the governing principle that *all* rules of *all* sports should be enforced less rigorously toward the end of contests? Presumably not. Few proponents of temporal variance, I'd wager, would contend that pitchers should be awarded an extra inch or so around the plate in the ninth inning, or that a last-second touchdown pass should be called good if the receiver was only a little bit out of bounds. So even if categorical temporal invariance is too rigid, the contours and bases of optimal temporal variance remain to be argued for. That is the task of this paper.

Or, I should rather say, that is this paper's surface agenda.

While seeking a deeper understanding of familiar sporting practices is a worthwhile project all on its own, this investigation is in service of a greater ambition. Jurisprudes have long drawn on games and sports for illumination.[5] Rawls, for example, famously drew on

[4] J. Everett Prewitt, Letter to the Editor, SPORTS ILLUSTRATED, Oct. 12, 2009, http://sportsillustrated.cnn.com/vault/article/magazine/MAG1161011/index.htm.

[5] Just what games and sports are, and whether the latter is a proper subclass of the former, are questions that I do not explore in this essay. Briefly, the Wittgensteinian teaching that games cannot be defined in terms of necessary and sufficient conditions was challenged some decades ago in a short, little-noticed book—BERNARD SUITS, THE GRASSHOPPER: GAMES, LIFE AND UTOPIA (1978)—that has recently won praise from such philosophers as Thomas Hurka

baseball to exemplify his practice conception of rules.[6] Hart invoked chess and cricket to illustrate both the internal aspect of rules and the difference between primary and secondary rules.[7] Dworkin also turned to chess to demonstrate the notion of constructive interpretation.[8] Examples like these could be multiplied with ease. Yet, jurisprudential attention to sports and games is decidedly ad hoc. I am unaware of any sustained or systematic investigation into the insights that formal sports and municipal legal systems might offer up for students of the other.[9]

The lack of sustained jurisprudential attention to games and, especially, sports should surprise, for sports leagues do, rather plainly, constitute distinct legal systems. This should be apparent on the surface things—at least to non-Americans. While the American sports scene is dominated by three home-grown team sports—baseball, football, and basketball—all of which are governed by official "rule books," the most popular global team sports like soccer, cricket, and rugby (both league and union) are all formally governed by "laws," not "rules." But the law-ness of sports systems is not merely superficial, for they exhibit such essential institutional features as legislatures, adjudicators, and the union of primary and secondary rules. We might reasonably have expected a discipline of "sports and law" to have arisen as a region of study belonging either to comparative law or to special jurisprudence.

To be sure, philosophy has spawned a subfield denominated philosophy of sport, actively represented by at least two societies (the International Association for the Philosophy of Sport, and the British Philosophy of Sport Association) that publish two specialty journals (*Journal of the Philosophy of Sport*, and *Sport, Ethics, and Philosophy*). But the integrity of this

---

and Simon Blackburn. I am skeptical that Suits's effort succeeds. (For a trenchant criticism see Norman Geras, *Games and Meaning, in* STEPHEN DE WIJZE ET AL., EDS, HILLEL STEINER AND THE ANATOMY OF JUSTICE 185 (2009).) Be that as it may, I think John Tassioulas clearly right, as against Hurka and Suits, in insisting that not all sports are games. *See* Hurka & Tassioulas, *Games and the Good,* 80 SUPP. PROC. ARISTOTELIAN SOC'Y 217 (2006).

[6] John Rawls, *Two Concepts of Rules*, 64 PHIL. REV. 3, 25 (1955).

[7] H. L. A. HART, THE CONCEPT OF LAW 89 (2d ed. 1994)

[8] RONALD DWORKIN, LAW'S EMPIRE 136–38 (1986); *see also* DWORKIN, TAKING RIGHTS SERIOUSLY 101–105 (1977).

[9] Indeed, most legal writing on sports that does not pertain to sports law is intended more to entertain than to edify, the best known example of the genre being Aside, *The Common Law Origins of the Infield Fly Rule,* 123 U. PA. L. REV. 1474 (1975). However, John Roberts's proposed analogy, offered during his confirmation hearings, between judging and umpiring, has recently provoked interesting jurisprudential writing—most of it (rightly) critical. *See, e.g.,* Aaron S.J. Zelinsky, *The Justice as Commissioner: Benching the Judge-Umpire Analogy,* 119 YALE L.J. ONLINE 113 (2010); Neil S. Siegel, *Umpires at Bat: On Integration and Legitimation,* 24 CONST. COMM. 701 (2007).

discipline is unclear—one resource proclaims, without apparent irony, that it draws on aesthetics, epistemology, ethics, logic, metaphysics, philosophy of education, philosophy of law, philosophy of mind, philosophy of rules, philosophy of science, and social and political philosophy[10]—inviting a reasonable worry that philosophy of sport might be about as coherent a field as "the law of the horse."[11] Moreover, even insofar as philosophy of sport encompasses the philosophy of law and "the philosophy of rules," it has not apparently drawn interest from scholars actively engaged in legal philosophy.

I think that is unfortunate. Legal theorists, after all, might be thought to be the experts in the problematics of rule-governed social institutions. The grander ambition of this paper, accordingly, is to help spur the growth of sports and law, or of the philosophy of sports and law, as fields worthy of more concerted theoretical attention. We might even say that this paper does double duty as a manifesto of sorts—not for slack, but for an enlarged program of jurisprudential inquiry.

Because this goal must strike some readers, and not only the more sports-phobic, as quixotic, perhaps I should note just some of the ways in which, as formal rule-governed practices, sports and law often pursue similar goals and confront many of the same challenges. For example, each domain must decide: to what extent to guide conduct by "formal" written norms as opposed to "informal" social norms, and, if the former, by rules or by standards; when to delegate discretion to the adjudicators (judges, juries, referees), and how best to constrain it; how to respond to the problem of epistemic uncertainty; whether to provide a right of appeal from unfavorable decisions and, if so, how to structure appellate review; how to conceptualize, deter, and sanction "cheating"; how to identify and rectify the gaps that inevitably arise between "the law in the books" and "the law in action"; when to tolerate ties and how to resolve them when they should not be tolerated; how to analyze and craft optimal sanctions; and so on, and so forth.

Finally, it's not just that (municipal) legal systems and sports systems confront similar challenges. There are reasons to believe that jurisprudential attention to the sporting domain

---

[10] LEON CULBERTSON ET AL., HIGHER EDUC. ACAD., RESOURCE GUIDE TO THE PHILOSOPHY OF SPORT AND ETHICS OF SPORT 1 (2008), http://www.heacademy.ac.uk/assets/hlst/documents/resources/philosophy_ethics_sport.pdf.

[11] *See* Frank H. Easterbrook, *Cyberspace and the Law of the Horse*, 1996 U. CHI. LEGAL F. 207.

is particularly likely to contribute to our understanding of the phenomena and dynamics shared in common. First, because the rules and practices of sports have long been viewed as unworthy of serious philosophical and theoretical investigation, even low-hanging fruit has yet to be harvested. Second, as we will see, sports supply a vast range of examples for the generation of hypotheses and against which to test our theories. And third, our judgments and intuitions about certain practices—such as, to take the present topic, the propriety of context-variant enforcement of rules—are less likely in the sports courts than in the courts of law to be colored or tainted by possibly distracting substantive value commitments and preferences.

In short, sporting systems, though rarely explored with seriousness by legal theorists and comparative lawyers, comprise a worthy object of legal-theoretical study. This essay is an extended illustration of that worth.

I. Preliminaries and the Plan of Attack

My broader ambition notwithstanding, our narrow subject remains whether, and under what circumstances, rules of sports should be enforced with greater laxity toward the end of close contests. Before tackling that question, a few caveats.

First, I mean this "should" in a legal, not a moral, sense. We have noted the possibility that the lineswoman should have called a foot fault on Serena Williams just so long as she genuinely believed that Williams's foot touched or went over the baseline before she hit the ball. Those who think otherwise could reach the contrary conclusion via varied routes. One possibility is to go outside the legal system. A proponent of this approach would agree that the rules of tennis concerning foot faults should be crafted and understood to be temporally invariant, but argue that "moral" or "all-things-considered" reasons dictated that the lineswoman should have refused to enforce the rule nonetheless. In directing our attention to the legal *should*, I mean to put that possibility aside. I am interested in the question of when the legal regime (here, the legal regime that makes up tennis) should itself provide for temporally variant enforcement of foot faults.

Second, I reject the conventionalist answer to this question. Some people think that what I claim to be a puzzle of genuine legal-philosophical interest is no puzzle at all. They say

that the answer is that it all depends on the "norms," "customs," or "conventions" of the sport in question—a matter about which they may cheerfully admit to have no expert knowledge. But this won't do. For we are seeking not simply a report of existing practices, but an account of what the practices should be. If, for example, it is the custom in some sport to afford competitors greater slack in certain game situations, the proponents of that custom should believe that it is backed by good reasons, and should hope to grasp what those good reasons are.

This is not to say that sport-specific norms and customs are irrelevant. Golf has a very different internal morality or ethos than does baseball.[12] It might well be that these differences properly bear on optimal practices regarding temporal variance. All I mean is that we ought not to assume that the existing practices concerning temporal variance are conclusive regarding what those practices should be, all things considered, even taking due account of relevant sport-specific features.

Third, when investigating the legal should, I intend at this early stage of the analysis to be agnostic among the variety of ways that an affirmative answer could be operationalized. Here, for instance, are five ways that temporal variance for foot faults could be realized: (1) the rules should be drafted expressly to specify the contexts in which touching the line or court does not constitute a foot fault; (2) the rules should be drafted expressly to specify the contexts in which touching the line or court, while an infraction, is unenforceable; (3) the rules should be drafted in a manner that does not expressly rule out context-variance, and should be interpreted and enforced in a context-variant manner; (4) the rules should be drafted and interpreted to confer discretion on the linesperson or umpire to adjudge infractions in a

[12] More noteworthy than golf's requirement that participants call infractions on themselves is the fact that players routinely do—even when the infraction was witnessed by no one else and when it conferred no competitive benefit. The legendary Bobby Jones, competing in the 1925 U.S. Open, conveyed the extent to which golfers have internalized their sport's code of honor when he assessed himself a 1-stroke penalty after he alone saw his ball move a fraction of an inch when he addressed it in the rough. Applauded afterwards for his integrity, Jones would have none of it. "You might as well praise me for not robbing banks," he protested. Baseball, in contrast, is widely understood, even glorified, as a game of cheating and deception—from spitballs and sign stealing to the hidden-ball trick. As Chicago Cubs President Andy MacPhail noted when his star slugger, Sammy Sosa, was caught with a corked bat, "There is a culture of deception in this game. It's been in this game for 100 years. I do not look at this in terms of ethics. It's the culture of the game." *See generally* JASON TURBOW & MICHAEL DUCA, THE BASEBALL CODES (2010).

context-variant manner; or (5) the rules should be drafted and interpreted in a manner that denies context-variance to adjudicating the fact of infraction, but confers discretion on the linesperson or umpire whether to enforce the prescribed penalty for the infraction. I am not at the outset distinguishing among these (or other) routes to temporal variance. Rather, the question of whether foot faults should be enforced in a temporally variant manner is essentially whether any one of these (or other roughly similar) propositions is true. I interpret temporal invariance to be committed to the proposition that none is true. If we conclude in favor of temporal variance, we can then explore whether the arguments in favor of temporal variance also bias in favor of one or another ways to implement it. Put slightly differently, assuming arguendo that foot fault penalties should be enforced with temporal variance (and therefore that Williams should not have been penalized), I am not presently focused on which agent of the tennis system—the gamewrights, the line judge, the chair umpire, etc.—should have done differently than they, she, or he did.

Fourth, my analysis is *pro tanto*, not conclusive. Figuring out whether a practice of temporal variance is optimal for any given sport, all things considered, is devilishly hard—too many considerations enter the scene, implicating too many contestable empirical and evaluative premises. But perhaps we should content ourselves with a more modest ambition. Perhaps it is enough, at least at this early stage, to try merely to figure out whether "sense can be made" of such a practice. Instead of trying to furnish a conclusive argument about just what the optimal practices should be, I will consider my efforts a success to the extent this essay explains how temporal variance could sensibly be—what can plausibly be said for it.

So much for preliminaries. Here is a short overview of the remainder of this essay.

Part II begins, not with tennis, but with other sports in which it has been alleged that a practice of temporal-variance is more secure—sports like football, hockey, and basketball most particularly. In each, whistles for minor physical contact toward the end of tight contests predictably elicit a cry from the stands: "Let 'em play!" Though the plea is familiar, its rationale is not obvious. To be sure, the tighter the rules are enforced, the less physical contact there will be. And reasonable fans and participants may reasonably disagree about the level of physicality that makes the sport in question the best it can be. But however we—or the

respective leagues—may answer that question, it is not self-evident why the optimal degree of laxity should be any different in crunch time during an NBA game, or throughout the NHL playoffs, than at any other time. It is not obvious, in other words, what can be said for "letting them play" *at this particular time* different in character or force from what can be said *generally* for "letting them play."

No, it is not obvious. Still, this is a good place to start. I'm skeptical that many tennis fans could assert with confidence that tennis officials allow players to get away with a little more foot faulting toward the end of close matches than earlier. Maybe they do, but foot faults just aren't called enough at any time to permit those without intimate knowledge of the sport to be sure what the patterns of enforcement are. Things are different with basketball. That basketball referees respect some measure of temporal variance seems clear enough to many hoops fans.[13] Just maybe that's because the case for temporal variance in basketball is unusually clear. If we can explain and justify slack in the calling of basketball fouls, we might be in stronger position to assess whether temporal variance makes sense in tennis too.

Unfortunately, whether the analysis that I end up offering in support of temporal variance in basketball and football applies to foot faults is uncertain. Part III explains why and proposes an alternative analysis that can help explain temporal variance in that context too. A short conclusion follows.

II. Swallowing the Whistle: Temporal Variance in the Enforcement of Fouls

I'm with those who believe that fouls are frequently called less strictly at the end of close games than otherwise. Even if true, this is hard to establish to the skeptical, for if this is the rule it is an example of "the rule in action," not "the rule in the books." Those portions of

[13] In the first round of the 2010 NCAA men's basketball tournament, a referee called a lane violation against a player for New Mexico State with 18.6 seconds remaining and the Aggies down to Michigan State 69-67. The call gave the Spartans another free throw, which they made for a 3-point lead that they didn't relinquish. Most commentators were withering in their criticism of the call. "You can't have that be a deciding call in a close NCAA tournament game," protested one. "That's a horrendous, horrendous call," objected another. "Lane violations happen on a majority of free throws and are almost never whistled. To call one with 18 seconds left in a two-point NCAA tournament game is unconscionable." Steve Schrader, *Critics: Refs Blew it on Lane Violation in Michigan State Game*, Detroit Free Press, Mar. 21, 2010, http://www.freep.com/article/20100321/SPORTS08/3210401/1055/Sports07/Critics-Refs-blew-it-on-lane-violation-in-Michigan-State-game

the official NBA rules that define and proscribe personal fouls do not provide that the kind or amount of contact required to constitute a foul varies according to contest-contextual features.

The most general statement of personal fouls provides that: "A player shall not hold, push, charge into, impede the progress of an opponent by extending a hand, forearm, leg or knee." More particularly, "[c]ontact initiated by the defensive player guarding a player with the ball is not legal," although "[i]ncidental contact with the hand against an offensive player shall be ignored if it does not affect the player's speed, quickness, balance and/or rhythm."[14]

Contrast this formal invariance with the NBA's remarkable varying standard of proof. As a comment on the administration and application of the rules directs: "there are times during a game where 'degrees of certainty' are necessary to determine a foul during physical contact. This practice may be necessary throughout the game with a higher degree implemented during impact times when the intensity is risen, especially nearing the end of a game."[15] Notice, then, that the formal rules that specify what constitutes a foul are context-invariant, but the standards of proof that determine whether a particular action will be adjudged to be of the forbidden type are context-variant.

An example might help make the contrast clearer. Suppose Referee is very confident that a defensive player, say the Spurs' Tony Parker, makes incidental contact with his hand against an offensive player, say the Miami Heat's Dwyane Wade. But Referee is very uncertain whether that contact affected Wade's speed, quickness, balance or rhythm. The "degree of certainty" rule seems to dictate something like the following: if Referee believes that it is only modestly more likely than not that Parker's contact affected Wade's speed, (1) he should call a foul through most of the game; and (2) he should not call a foul with the game tied in the last minute. In contrast, as far as the formal rules provide, were Referee fairly confident that Parker's contact did affect Wade's speed (by slowing him), but only to a slight degree, then he should enforce the foul no matter when it occurred. One way to think about our task is as an attempt to figure out what could be said for bringing greater symmetry to the rules—for allowing the referee the same discretion in determining how much contact (or how much

[14] Official Rules of the NBA, Rule 12B, Section I.
[15] Official Rules of the NBA, Comments on the Rules: I. Guides for Administration and Application of the Rules, *available at* http://www.nba.com/analysis/rules_index.html.

impedence) constitutes a foul at different points during the game as he is granted in determining how confident he must be that the forbidden amount of contact was committed.

One answer invokes essentially aesthetic considerations: the referee's whistle disrupts play, thereby reducing spectators' enjoyment of the action. To be sure, disruption of play almost always incurs an aesthetic cost. But inasmuch as dramatic tension builds during crunch time, disruption of play during this time is especially costly.

There is something to this justification for temporal variance. It would seem to apply, though, only when play would continue uninterrupted but for the calling of a foul. However in some sports that arguably respect temporal variance play stops either way. It appears to me, for example, that football officials are often a little more reluctant to call defensive pass interference during crunch time even though an incompletion stops play just as surely as does a penalty flag.[16] Because an aesthetic or dramatic preference that play not be disrupted would seem not to explain or justify temporal variance in the calling of fouls and the enforcement of penalties across the board, it might not provide the whole story even in basketball. So without denying that appreciation for dramatic excitement can help explain why officials should give the competitors somewhat greater slack during moments of high drama, it behooves us to explore the possibility of an alternative account too.

The alternative account that I'll offer has two core components. First, expanding on the wisdom at the heart of the saying "no harm, no foul," I argue (in Sections A through C) that, insofar as penalties are designed to serve a compensatory or restitutionary function, we have reason not to impose them when they would work substantial overcompensation. Second, starting from the straightforward idea that the competitive cost imposed by any given

---

[16] And not only to me. Here is *Sports Illustrated*'s lead NFL columnist, Peter King, explaining his naming Browns cornerback Hank Poteat "goat of the week":

> with Cleveland holding a 37-31 lead and no time left on the clock in the fourth quarter, Detroit quarterback Stafford let fly with a rainbow to the end zone and Poteat tackled Calvin Johnson with the ball in the air. *If Poteat had jostled Johnson, there's little chance a flag would have been thrown*. But a full-scale body slam to the ground ... That has to be called. Pass interference. With the extra play, Detroit threw a touchdown pass to win it. On the goat scale, Poteat's play ranks about as high as you can go.

Peter King, *Monday Morning QB*, SI.COM, Nov. 23, 2009, http://sportsillustrated.cnn.com/2009/writers/peter_king/11/22/Week11/3.html (emphasis added).

outcome-affecting contest event—e.g., a score, an infraction of the rules that confers a competitive benefit on the rule-breaker, a penalty imposed by an official in response to an infraction, etc.—is not constant, but context-variant, I draw attention to one contextual factor of particular significance: time (or functional equivalents). In particular, I argue (this is Sections D through F) that the competitive impact of an event occurring during a close contest varies in inverse proportion to the distance remaining to the contest's completion.[17] If these two claims are correct, then it follows that a penalty of nominally constant magnitude that it is optimal to impose early in a contest may become suboptimal to impose later in that same contest.

A. On the Saying, "No harm, no foul"

Take a step back and consider the familiar saying "no harm, no foul." (NHNF) We hear it, and use it, frequently. But what, exactly, does it mean?

I'm not the first to ask this question. "Curious" posed it a few years ago on Yahoo Answers, eliciting this "best" response: "No one was hurt, nothing is wrong." Or, as another reply offered: "it means that if noone [was] hurt in what you were doing, then what you have done is justified."[18] Another website, "UsingEnglish.com," elaborates: "There's no problem when no harm or damage is done, such as the time my sister-in-law stole the name we'd chosen for a boy and we both ended up having girls."[19]

This is bad ethics. Take the case of the thieving sister-in-law. Admittedly, I'm not entirely certain what is involved in stealing a name: despite what a surprising number of parents (perhaps teenage mothers especially) appear to believe, one *can* give a child a name

---

[17] Many sports (e.g., basketball, football, hockey, soccer) are timed. In such sports, distance to contest completion is temporal. Unclocked sports (e.g., baseball, tennis, golf, volleyball) use mechanisms other than the passage of time to determine when a contest ends. In them, distance to completion must be measured in other units—in baseball, the units are outs that a team may incur or must secure; in tennis and volleyball, they are games that one must win or may lose; in golf, the units are holes. So these other means of measuring progress toward contest completion—outs, games, holes, etc.—are what I mean by "functional equivalents" of time. For simplicity of exposition, I will speak about time and temporality but with the understanding that I mean to invoke its functional equivalents when the context requires.

[18] Postings of minaleri & lostlight21 to http://answers.yahoo.com/question/index?qid=20070518143109AAx6BR1 (May 18, 2007).

[19] Idiom Definition: No Harm No Foul, http://www.usingenglish.com/reference/idioms/no+harm,+no+foul.html (last modified Dec. 18, 2009).

already in use.[20] Still, the scenario I imagine goes like this. One pregnant woman learns of the name that her pregnant sister (or sister-in-law) intends to give her baby, if it is a boy, and forms the intention to give it to her own baby, if a son, knowing that doing so would make the name no longer attractive, or significantly less attractive, to the couple who first had the idea—that is, who had the idea before she did. To be sure, that both women bore girls softens the sting of the betrayal. But the notion that the sister-in-law did nothing wrong—or, in any event, that "there's no problem" with her conduct—seems plainly mistaken. We are told, after all, that she "stole" something (a name). That would seem to be wrongful if true. That the wrong proved harmless is fortunate, but not an erasure of the wrong itself.

The answers proposed by the Yahoo responders, though admirably concise, are even further off base. One can run afoul of standards of rightful or permissible behavior even without causing harm, which is why almost nobody thinks criminal punishment for unsuccessful complete attempts is unjust or morally impermissible (assuming that criminal punishment is not generally unjust or impermissible). So if "no harm, no foul" has any sensible meaning, we still haven't identified just what it is.

True story: some evenings ago my wife and I returned home to find our house empty, but the front door unlocked. Our kids' babysitter had taken them to the park and had forgotten to lock up. My wife remarked upon it when he returned. "Gosh," he said embarrassedly. "You're right. I just forgot. I'm really sorry." "Well, it's okay," she assured him. "No harm, no foul."

That, I think, is a sound usage. But my wife wasn't saying that our babysitter's lapse was "justified," nor was she denying that he had, in fact, done something wrong—an unintentional wrong to be sure, but a wrong nonetheless. At least in my wife's deployment of the phrase, "no harm, no foul" is a performative—namely, an acceptance of an apology. Thus does another

[20] Recall from *Freakonomics* the remarkable statistic that nearly 30 percent of black girls born each year in California are given a name that is given to no other California baby born that year. STEVEN D. LEVITT & STEPHEN J. DUBNER, FREAKONOMICS: A ROGUE ECONOMIST EXPLORES THE HIDDEN SIDE OF EVERYTHING 184 (2005).

web dictionary define the phrase: "no problem, it's cool. Usually used in response to someone's apology to indicate acceptance."[21]

Insofar as this is correct, it has an important implication for our question. To see why, we should distinguish two things that are often conflated: accepting an apology and demurring to it.

Sometimes we are causally responsible for bad states of affairs even when not to blame. Take this common example: D is driving down a residential street at an appropriate rate of speed and with utmost care. C, a young child, darts out from behind a parked car into the path of D's minivan. With no time to react, D drives into C, killing him. D has a relationship to the event that P, a pedestrian down the block, lacks: D caused the death of a child. And yet, on these facts, he is not to blame. Moreover, on the dominant view, he hasn't even committed a wrong. [22] It is tragic all around—tragic for C, for C's family, and for D himself—yet, for all that, a quintessentially blameless accident. Of course, though, D's blamelessness doesn't let him entirely off the hook. Bernard Williams famously contended that D ought to experience agent-regret.[23] At a minimum, he has a duty to express remorse: "I'm so terribly sorry."

It seems to me that C's family, if they understand the relevant basic facts, ought to, let us say, "let D off the hook." But there are different ways to do so. (Sadly, in this case, "no harm, no foul" is not among them.) Imagine this choice: "Thank you, we accept your apology." If you're D, you might think this not quite apt. You might feel, whether or not you choose to voice it, that by "I'm sorry" you didn't mean "I apologize"—at least not in a robust or unqualified sense. You felt deep regret for the outcome and remorse for your causal role, but were not intending to own a wrong. The better response of C's family would have been to acknowledge that fact. "It wasn't your fault" would have been more apt than "We accept your apology."[24]

---

[21] Urban Dictionary: No Harm No Foul, http://www.urbandictionary.com/define.php?term=no%20harm%20no%20foul (posted Feb. 22, 2005).

[22] Dominant, but contested. For variants on the opposing view see, e.g., MATTHEW H. KRAMER, WHERE LAW AND MORALITY MEET (2004) (arguing for strict moral liability); John Gardner, *The Wrongdoing that Gets Results,* 18 Phil. Perspectives 53 (2004) (arguing for moral duties to succeed).

[23] BERNARD WILLIAMS, *Ethical Consistency*, *in* PROBLEMS OF THE SELF 166–86 (1973).

[24] In a context such as this, "it wasn't your fault" is the idiomatic way to deny, not merely that the actor is blameworthy, but also that he did anything wrong. I believe such an expression appropriate. As Matt Kramer

The point of this sad little story is that accepting an apology is not the only alternative to rejecting it. Rather, there are at least three responses one can make to an apology: rejection, acceptance, or demurrer. This third option is to say that the apologizer has done nothing for which the need to apologize arises.[25] The second option is less indulgent, for it avows precisely what the third denies: that the actor *has* committed a wrong for which an apology is required. So insofar as "no harm, no foul" serves to accept an apology, then its force or upshot is not to *deny* the commission of a foul, but to *affirm* it.

This understanding of the saying comports with its original usage. Longtime Lakers announcer Chick Hearn coined the expression in the 1960s to express the idea that referees should not call minor fouls that do not interfere with the flow of play. The definition of the adage offered by Wiktionary is consistent with its use by both Hearn and my wife: "Encapsulation of the idea that although technically a breach of some code or law may have occurred there is no need for punishment . . . or retribution if no actual damage occurred."[26] It offers an example: "He parked in my space but as I was away at the time: no harm, no foul." In contexts such as these, "no harm, no foul" is a slight misnomer. The underlying idea would be rendered more accurately, if less gracefully, as *no harm, no penalty, notwithstanding foul.* Call this reformulation NHNP.

B. The puzzle of "no harm, no penalty."

Now we encounter a modest puzzle. Some norms and rules can be violated only by the causing of harm or injury, whereas other rules can be violated by proscribed conduct all alone, regardless of whether that conduct causes any further bad state of affairs. As a very rough generalization (albeit one subject to many exceptions), the law of torts (largely designed to

---

rightly emphasized to me in private correspondence, his view accepts that C's family should acknowledge D's faultlessness or blamelessness, but not his purported lack of wrongdoing.

[25] I am somewhat simplifying a yet more nuanced moral landscape. It might be that D does have a duty to apologize, as P of course would not, but that C's family has a duty to affirm, in response to the apology, that D had no such duty. Morality and etiquette are tightly intertwined here, for while our fundamental obligation is to respect and express respect for other persons, we often have no better way to satisfy that obligation than to conform to the complex social rituals designed to pattern respectful behavior.

[26] Wiktionary: No Harm No Foul, http://en.wiktionary.org/wiki/no_harm,_no_foul (last modified Dec. 16, 2009). Omitted by the ellipses is the word "apology." I think that Wiktionary is wrong about that. In paradigmatic cases where no harm, no foul reasoning applies, there *is* a need for an apology, though the apology should be accepted. That is why it was my wife, not our babysitter, who uttered "no harm, no foul," and why she did so *after* he had apologized.

ensure compensation for injury) requires harm for the commission of a foul, the criminal law (principally designed to deter commission of serious wrongs and to inflict retribution for blameworthy wrongdoing) does not. Sports rules are similarly varied: some are defined such that their commission requires a proscribed result, most aren't. Examples appear in the margin.[27]

"No harm, no foul" is often invoked to urge that a penalty not be imposed even when the relevant rules provide that harm is not required for the foul or for the consequent penalty, which is what Chick Hearn meant to express and why Wiktionary has things right in emphasizing that, frequently, the adage is offered in response to "a breach of some code or law." The puzzle, then, is this: given that rulemakers know how to draft rules so that their violation does or does not require harm, and know how to specify that a harmless violation should incur no penalty, why would the non-realization of harm be thought to warrant non-imposition of a penalty even in cases where the rule does not require harm?

---

[27] One sport that resolutely embraces the principle "foul, regardless of harm" is basketball: the victim of a shooting foul goes to the charity stripe even if his shot had gone in. Golf is similar. For example, a player is disqualified for signing a scorecard that reports a lower score for any hole than actually taken. (USGA Rules of Golf, Rule 6-6.d.) Imagine the player who scores a 3 on hole 12 and a 4 on hole 13, but who accidentally switches the scores when recording them. The final recorded score is correct, but the player is nonetheless disqualified.

In football, in contrast, most infractions incur penalties regardless of whether the infraction affected the play. For example, when a long run from scrimmage (or a long return of a punt or kickoff) is called back because of a hold or an illegal block by the offense (or the receiving team), it is not uncommon for the announcer to observe, in commiseration with the penalized team, that the infraction was particularly unfortunate because it occurred at a place on the field where the player held or blocked could not possibly have had an impact on the play. The rule for pass interference is different: "Contact that would normally be considered pass interference [is not pass interference if] the pass is clearly uncatchable by the involved players." (Rule 8, section 2, article 5(c).)

In soccer, the offsides rule is much like the uncatchable rule in football: a player who is in an "offside position" does not commit an "offence" if he is not "involved in active play" by interfering with the play or with an opponent, or by "gaining an advantage by being in that position." (FIFA Laws of the Game, Law 11.) But what about the player who simulates an injury with the intent of deceiving the referee? This is an offence for which a yellow card must be awarded regardless of whether the player has successfully deceived anyone. (Law 12, Decision 5.) In table soccer (aka foosball), "spinning the rods is illegal" (International Table Soccer Federation Rule 15) but "spinning of a rod which does not advance and/or strike the ball does not constitute an illegal spin." (Rule 15.2)

Most of the sports rules that make harm matter actually take the form NHNF, not NHNP. For present purposes, this difference is not material. The present point is that rule makers know how to specify that harm is a necessary condition either for a finding of violation or for the imposition of a penalty. Therefore, their failure to avail themselves of either option suggests that a given rule of conduct can be violated, and its violation penalized, even in the absence of a harmful consequence.

The question is sufficiently important that I'll risk belaboring it. If in a *particular* case no penalty should be imposed because no harm was caused notwithstanding that the conduct violated a rule that is not defined in terms of causing of harm, then why doesn't the reason for not imposing the penalty serve as well as a reason for reframing the rule to require harm-causation in *all* cases? Put another way, is there an argument for not imposing a penalty on the commission of a foul in a particular case because, in that case, the foul caused no harm that allows for the possibility that the penalty should be imposed in other cases just because of the foul and without regard for whether harm was caused? The plausibility of "no harm, no penalty, notwithstanding foul" depends on explaining how this no-harm-causing act-token of a proscribed act-type differs from other no-harm-causing act-tokens of that same proscribed act-type.

Briefly and quite generally, I think the answer (or part of it) is this. Were a penalty imposed solely for compensatory or restitutionary purposes then we would have no reason to enforce it when particular fouls end up being costless. Furthermore, we have affirmative reasons *not* to enforce it. Among other things, such penalties disrupt the flow of the contest and they handicap a competitor who has imposed no cost on his opponent. But of course penalties usually serve other functions too—most importantly, deterrence. This is why penalties authorized for violations of act-types that are proscribed because of their tendency to cause harm are usually sensibly enforced even in response to tokens of those act-types that happen not to cause any harm.

Usually, but not necessarily always. On this account, non-enforcement might be warranted if, for context-specific reasons, enforcement of a penalty in a particular case would be unusually costly to the rule-breaker (or to other interests), or if non-enforcement of the penalty on this occasion would weaken the deterrent force of the rule to an unusually small degree.

Let's recap. We started by asking what "no harm, no foul" means. We considered and rejected one oft-suggested answer: that an action is necessarily justified or not wrong if it fails to produce harm. Instead, I proposed that it is often used to accept an apology in circumstances when somebody did do something wrong but, happily, nothing bad came of it. We saw that

this usage gives the lie to the expression because it means not that no foul has occurred but that no bad consequences should be imposed even though a foul *did* occur. And we saw that this formulation is curious. Because rules are sometimes crafted to make occurrence of harm a necessary condition for the foul and for any prescribed consequences, when a rule is crafted otherwise, the apparent implication is that any prescribed penalty should attach regardless of whether harm has occurred. Such a practice, despite a possible whiff of unfairness, can be easily defended on the ground that imposing a penalty regardless of harm improves deterrence of risky conduct in the future. Finally, we said that it might make sense to refrain from imposing a penalty even of such rules if, in a given case, no harm has occurred *and* either enforcement of a penalty in a particular case would be unusually costly to the rule-breaker (or to other interests), or if non-enforcement of the penalty on this occasion would weaken the deterrent force of the rule to an unusually small degree.

If this is right, then it should be apparent what more we need to explain and justify temporal variance in the enforcement of sports penalties. We need to establish that, at crunch time, penalizing harmless fouls would be unusual in one of the respects just mentioned. Before exploring whether that might be so, we should examine one small wrinkle. We have translated "no harm, no foul" into "no harm, no penalty, notwithstanding foul." But what if the conduct doesn't cause *no* harm? What if it causes *some* harm, but very little?

C. From "no harm, no penalty" to "little harm, no penalty"—or not?

If prescribed penalties should sometimes not be enforced against the violator of a rule that is itself defined without regard to harm, when and because that particular violation caused *no* harm, does it also follow that there will be some occasions (albeit fewer) in which the penalty should not be enforced because the particular rule violation caused only *minor* harm? That is, if we are forced to choose one of two suboptimal alternatives—either (a) leaving a minor injury entirely uncompensated and thus allowing a rule-breaker to enjoy the benefits of his violation, or (b) substantially overcompensating the victim and thus shifting a cost upon the rule-breaker substantially in excess of the cost he would otherwise be allowed to impose on this victim—ought a sensible system ever to allow the injury to go unremedied?

One might think not. It is tempting to suppose that, as much as we might prefer perfect restitution, if the only options available to us are overpenalizing a rule-breaker or undercompensating his victim, justice or fairness demands that we always select the former. We might say that, by violating the rule, an actor has "assumed the risk" that he'd be subject to a disproportionately excessive penalty, or that he forfeited his claim against a disproportionate penalty. Notions like guilt, fault, and innocence are surely relevant to our choice between suboptimal alternatives. The question is whether they are always decisive.

The short answer is that they aren't—a proposition hard to conclusively establish, but easy enough to illustrate with well-settled legal practices that require victims of wrongdoing to simply lump it when the harm they have actually incurred is too slight.

One well-known example is contract law's material breach (or "substantial performance") doctrine, familiar to generations of law students from the leading case of *Jacob & Youngs v. Kent*.[28] Jacob & Youngs contracted, in 1913, to build a Long Island home for a property owner at a total cost just north of $77,000. The owner specified that all pipe used in the building must be manufactured by the Reading Manufacturing Company. In the event, and probably inadvertently, the contractor installed pipe manufactured by other companies, though generally understood to be of quality and price equivalent to that made by Reading. When, after the house had been substantially completed, the owner discovered the discrepancy, he refused to pay the remaining balance of $3500 and instructed the contractor to remove the offending pipe and replace it with pipe that would conform to the contractual requirements—namely, Reading pipe. The New York high court, in an opinion by Benjamin Cardozo, held for the contractor.

There was no doubt that the contractor had breached his contract. And the usual remedy for breach was to pay money damages sufficient to put the victim of the breach in the position for which he had contracted. But not in this case, said the court. Because the breach was minor and full performance would have "meant the demolition at great expense of substantial parts of the completed structure,"[29] the court refused to order a remedy so

[28] 230 N.Y. 239 (1921).
[29] 230 N.Y. at 240-41.

"grievously out of proportion" to the home owner's injury.[30] Instead, it held that the contractor should be required only to pay the difference in market value between the installed pipe and that specified by contract—which was to say, on the facts of this case, nothing.

The "harmless error doctrine" governing appellate review reinforces the point—especially as applied to appeals from criminal defendants. Suppose an appellate court is persuaded that a criminal defendant's legal rights were violated at the trial that resulted in his conviction—say, the trial judge improperly commented on the defendant's decision, protected by the Fifth Amendment privilege against self-incrimination, not to testify on his own behalf; or admitted an out-of-court statements obtained from the defendant in violation of his right to counsel; or erroneously barred the defendant from representing himself. Possibly the most natural remedy would be to vacate the conviction and order a retrial. But trials are expensive and time-consuming. And because so many criminal trials are infected by minor errors, requiring retrial in all such cases would, in the aggregate, substantially increase the criminal courts' burdens, increasing delays across the board. Furthermore, things may have changed since the initial trial that make a conviction unreasonably more difficult—memories become hazier, a key witness might have died. For all these reasons, courts and legislatures have concluded for nearly a century and a half—starting with the English Judicature Act of 1873—that retrial should not be automatically ordered.

While the precise tests employed by the courts are complex and vary across jurisdictions, the common theme is simple. Except for a few constitutional violations that threaten core notions of procedural justice (prosecution in violation of an individual's right to be free from double jeopardy, for instance, or the flat denial of a defendant's right to counsel) constitutional violations at trial will not be remedied by vacation of the conviction and an order of retrial if the appellate court concludes that the legal error did not likely contribute to the jury verdict—if, in the language of the doctrine, the error was probably "harmless." Although many aspects of the doctrine are controversial, the basic idea that a just system of criminal law need not remedy all violations of trial rights is rarely contested. But the doctrine is misnamed. The inquiry does not truly identify errors that are wholly harmless to defendants: if nothing else, the

[30] 230 N.Y. at 244.

violation of a constitutional right is almost invariably injurious to the right-holder's dignity and justified sense of entitlement. Rather, much like the superficially dissimilar material breach doctrine, it serves to withhold remedies that are "grievously out of proportion" to the injury sustained.

The lesson is simple. Just as *no harm, no foul* (NHNF) is frequently more accurately rendered as *no harm, no penalty* (NHNP), NHNP entails as well *little harm, no penalty* (LHNP).

D. On the saying "It cost us the game."

In week six of the 2009 NFL season, the undefeated Minnesota Vikings hosted the 3-2 Baltimore Ravens in what turned out to be a thriller. Up by 17 with ten minutes to go, the Vikings looked like they were on their way to a blowout. But second-year QB Joe Flacco led the Ravens to three quick touchdowns to take a 31-30 lead with under four minutes remaining. The ageless Brett Favre responded, driving the Vikings to a go-ahead field goal just inside the two-minute mark. At the end of regulation, Ravens kicker Steven Hauschka missed wide on a 44-yard field goal attempt that would have given the Ravens the victory. Not surprisingly, many fans bemoaned that the miss "cost us the game."[31] But Ravens running back Ray Rice, whose 194 yards from scrimmage were wasted in the loss, demurred: "We didn't lose that game because of Hauschka's miss. If we start fast and put points on the board, our defense starts fast, I think the game is a totally different outcome."[32]

This is a common routine across team sports. After some late-game mishap leads to a loss—a dropped pass or blown coverage, a pair of missed free throws, a base-running error, or a referee's bad call—a large number of fans are sure to complain that the mistake "cost us the game." Other fans, players or coaches, will disagree. No, they'll protest, *that* play didn't cost us the game. Had we played better in other facets of the game—had we been able to convert on earlier trips into the red zone, or had we left fewer men on base, or had we capitalized on our power plays—we wouldn't have been in that situation in the first place.

---

[31] E.g, Posting of It's NO GOOD! to http://weblogs.baltimoresun.com/sports/ravens/blog/2009/10/post_4.html (Oct. 18, 2009);

[32] Clark Judge, *Vikings Barely Stay Perfect, Hold off Flacco, Ravens*, CBSSPORTS.COM, Oct. 18, 2009, http://www.cbssports.com/nfl/gamecenter/recap/NFL_20091018_BAL@MIN.

The second assertion is surely right. Contrary to the adage, defeat has as many fathers as does victory. It can fairly be said about any number of plays that, had it gone differently, the loss would have been a win. But what about the first part of the protest? Does it follow from the fact that many players and coaches share responsibility for the loss that the particular poor play or bad decision did not cost the team the game? Was Ray Rice right that the Ravens "didn't lose that game because of Hauschka's miss"?

Perhaps it depends on what we mean by "because of." Had Hauschka not missed, he would have made it. (This assumes, of course, that the closest possible world to the actual is one in which the field goal is still attempted.) And had he made it, the Ravens would have won, 34-33. So Hauschka's miss was a but-for cause of the Ravens' loss: but for the miss, they would not have lost; they would have won. In this fairly straightforward sense, the Ravens *did* lose because of Hauschka's miss. But—and this is very likely what Rice really meant—the Ravens didn't lose *only* because of Hauschka's miss. There were many but-for causes of their loss.

So if it is true that Hauschka's miss was a but-for cause of the loss, but wasn't the only such cause, why focus on it? I think that Rice and others who say similar things are urging that we really ought not to. That we focus on it only because it is more salient to us, perhaps because we know how things would have turned out had he done otherwise, whereas, although there are many other but-for causes, we forget about them, or don't know which they were. Rice is claiming that Hauschka didn't cause any more damage than did any other Raven who missed a play in that contest.

That is the question to investigate. More precisely, of course, we're interested in the more general question that this particular query exemplifies. The general question is whether the costliness of detrimental contest events varies depending on time of contest. And here's a way to focus the question in a pretty clean way. Suppose two missed 44-yard FGs—the first with time expiring in the first quarter, the second with time expiring in the fourth, both with the kicking team down 10-9. Are the two misses equally costly, or was one more costly than the other? And if one was the more costly, which?

The more common intuition, I venture, is that the later miss is more costly than the earlier one. But many colleagues with whom I have spoken answer that the two are equally

costly. I will argue that the latter is more costly than the former. And, therefore, that there is a straightforward sense in which Hauschka's miss was especially costly precisely because of when it occurred. It's not just a matter of drama and atmospherics. Instead, generally speaking and all else equal, events have greater impact on the outcome of a game—good things contribute more to victory, and bad things are more costly—when they occur later in close contests. If that argument succeeds, we'll have the final piece of the explanation for slack in the calling of basketball fouls and similar infractions.

E. "Crunch time" and the varying magnitude of outcome-affecting events

How ought we to think about the costliness of outcome-affecting events? I propose that the competitive costs or benefits of *any* game action*, x,* can be conceptualized in terms of the action's impact on a given competitor's probability of victory. Assuming that the probability of victory ranges from 0 to 1, the competitive cost or benefit of an action, *x,* ranges from -1 to +1. Let us represent the impact of an action, *x,* on competitor A's probability of victory by $\phi_{X(A)}$. Now, let In-m stand for a token of infraction-type n committed by competitor m, and let Pn-m stand for a token of penalty-type n imposed on competitor m. Thus: $\phi_{In\text{-}A(B)}$ is the impact of a given infraction by A on B's probability of victory, and $\phi_{Pn\text{-}A(A)}$ is the impact on A's probability of victory of being assessed some given penalty. Lastly, because penalties for fouls in sports including basketball, football, and hockey are designed to deter, they must be (by and large) over-compensatory.[33] So: $|\phi_{Pn\text{-}A(A)}| > |\phi_{In\text{-}A(B)}|$.

[33] You might think that the common practice in basketball of intentionally fouling when behind late in games—when stopping the clock is more important than not giving up points—is a counterexample to the claim in text: the competitive benefit from fouling (and the competitive cost of *being fouled*) is greater than the competitive cost of free throws to one's opponent. If it weren't, teams wouldn't keep doing it.

In fact, though, the practice just shows either of two things. First, it might be that intentional fouling late in games *is not prohibited.* If you drive westward from Manhattan over the George Washington Bridge, you will be compelled to pay a toll on the far side. That toll is not a penalty and your driving into New Jersey is not prohibited. The toll, while no doubt unwelcome to you, is a price exacted for permitted conduct, not a penalty or sanction imposed for prohibited conduct. On the distinction, see generally Robert D. Cooter, *Prices and Sanctions,* 84 Colum. L. Rev. 1523 (1984). In much the same way, organized basketball might be permitting teams to intentionally foul late in the game for a price. And while that might seem absurd, something can be said for it. While nobody much likes the stretching out of games with one team fouling the other on each possession, to prohibit the practice would be to say that it's better to let a team with a lead pass and dribble out the clock. It's not crazy to think that would be worse.

But maybe you think that would not be worse. Maybe it would be better to prohibit fouling late in games even if it means that one team gives up any realistic chance to win. If fouls are always prohibited and never meant to be permitted-for-a-price, then we come upon the second possibility: the rules of basketball are not, in this

It should be apparent on little reflection that both $\phi_{In\text{-}A(B)}$ and $\phi_{Pn\text{-}A(A)}$ are context-variant to at least some extent. For example, they vary depending on the score differential at the time of the event: each is greater when the score is tied than when either team enjoys an insurmountable lead. The claim I want to add is that, holding all else constant, they are also temporally variant. This is for the simple reason that, holding closeness of contest constant, scores and missed scoring opportunities have greater impact on the outcome of the game the closer they occur toward the game's end.[34]

Here's an illustration: Suppose that the Lakers are called for a shooting foul on a missed basket by the Celtics in the first 20 seconds of a scoreless scheduled 48-minute game. Suppose that the shot missed the basket and backboard and sailed out of bounds. Absent the enforcement of a penalty, the Lakers would have been awarded the basketball and the score would have remained 0-0. Assuming that the teams were evenly matched and playing at a neutral site, the Lakers' probability of victory absent a penalty would have been, let us suppose, .5. The awarding of two free throws to the Celtics as a penalty for the Lakers' foul has an expected value of 1.5 points. The Lakers' probability of victory when down 1.5 with 47:40 remaining is, let us imagine, .495. The cost of the penalty to the Lakers is -0.05, a small cost indeed. Now assume the same facts, except that the foul is called and penalty assessed with time expiring and the Lakers ahead by 1 point. The Lakers' probability of victory absent the penalty would be exactly 1.0: the game would have ended at that moment with the Lakers

---

respect, well-designed. The league should increase the penalty for intentional fouling enough to substantially reduce the incidence of its occurrence.

Could it do so? Sure. A contrast between football and soccer is especially revealing here. In football, a cornerback who knows he has been beaten will often grab the wide receiver. This brings forth a call for either pass interference or defensive holding (depending on whether the ball is already in the air). But either penalty might be less costly to the defensive team than allowing the receiver to continue unmolested. In effect, the cornerback is treating the penalty as a price not as a sanction. And the fact that announcers are likely to congratulate the defensive back for his heads up play suggests that insiders to the sport believe such conduct is permitted-for-a-price, not prohibited-on-pain-of-penalty. Much the same situation arises in soccer when a defender playing the last line of defense is beaten by a ball handler. When the ball handler has a sufficiently clear run to the goal, the defender might well think it cost-effective to foul and pay the price of a free kick. Soccer, unlike American football, decided to resist the attempt by players to convert what had been intended as a prohibition into a permission. Its response was to make an intentional foul under such circumstances (what is often, if misleadlingly, termed a "professional foul") a red card offense.

[34] The closeness of a contest is a function, inter alia, of the score differential and the distance remaining to contest completion. (*See supra* note 17.) For a constant score differential of n (>0), the contest is closer at t1 than at t10. What I mean to keep constant is closeness of contest not score differential.

enjoying a razor-thin lead. The probability, post-penalty, that the Lakers will win is the probability that the Celtics will miss both free throws (.06) + the probability that the Celtics will hit only one of the two free throws and then lose in overtime (.38*.5), or 0.25. The cost of the penalty to the Lakers under these circumstances is a whopping -0.75.

If that is so, then we can see the reason for preferring that the penalty not be imposed for this particular infraction: *We want the outcome of athletic contests to depend (insofar as possible) upon the competitors' relative excellence in executing the particular athletic virtues that the sport is centrally designed to showcase and reward*, and a sanction of this magnitude would make the outcome too dependent on the less important (though not unimportant) ability to refrain from any bodily contact. Let me be very clear: what I have just said is not to claim that this latter excellence is something that, in the nature of things, no sport could wish most to valorize; I am claiming only that, in the sport of basketball as we know it, this particular excellence does not rank so highly among the excellences that we wish to feature and encourage.

Or consider another illustration, moving from basketball to football. The Cowboys are beating the Giants 21-20 with 1:00 remaining in the fourth quarter and the Giants facing fourth and ten from their own 40. Giants QB Eli Manning throws the ball 40 yards downfield, where a Cowboys' cornerback interferes modestly with the Giants' receiver. Even absent any interference, it would have been a very difficult catch—say, a .3 probability. The modest physical contact made a reception yet more difficult, but still possible—say, a .2 probability. Either way, had the receiver caught the ball, the reception would have taken him out of bounds at the Cowboys' 10-yard line. In the event, the pass falls incomplete. On these assumptions—stylized but not fanciful—the interference imposed a cost of -0.1 likelihood of reception. A reception at this field position and time, let us say, would have given the Giants a .96 probability of victory. An incompletion gives possession to the Cowboys, and thus leaves the Giants with a mere .01 probability of victory.

Should the foul be called and the penalty—awarding the Giants possession and a first down at the spot of the infraction—enforced? Many informed observers would say no, that during "crunch time," the referees should give the players a little more latitude for physical

contact.[35] Without fully resolving whether, on balance, this is so, the analysis to this point at least makes sense of why it might be. The contest-outcome cost of this defensive pass interference on the Giants ($\phi_{In\text{-}C(G)}$) is -0.095. The context-outcome cost of the penalty on the Cowboys ($\phi_{Pn\text{-}C(C)}$) is -0.95. When a penalty would have such a large expected impact on the game outcome and when the unpenalized infraction would not have an impact of roughly similar magnitude, we might think that the game goes better—it is fairer and more satisfying—by letting the play on the field stand.[36]

In short, the expected outcome-affecting magnitude of an outcome-affecting event is greater toward the end of a contest than at its start, holding closeness of contest constant, because there are fewer opportunities for the impact of that event to be countered.[37] Another

---

[35] *See supra* note 16.

[36] An obvious alternative to ignoring the infraction entirely, of course, would be to impose a more modest penalty. Possibly, the NFL makes things unnecessarily hard on itself by insisting that defensive pass interference must always be penalized at the spot of infraction, instead of allowing for more modest yardage mark-offs for minor interference. (The NCAA penalizes defensive pass interference at the spot of infraction only if within 15 yards from the line of scrimmage; if the infraction occurs farther downfield, it incurs a 15-yard penalty. 2009-10 NCAA Football Rule 7-3.) But it's unlikely that the NFL will find this proposal congenial, for in the long-running battle between "lumping" and "splitting" (categorizing heterogeneous phenomena into few broad classes that emphasize similarities or into many narrow classes those emphasize difference), the league's recent rescission of the 5-yard face masking penalty in favor of making all such infractions punishable by 15 yards suggests that it has cast its lot with the lumpers. And this isn't obviously crazy. Sensible system designers will frequently make available a more limited range of penalty options than might initially strike us as possible and desirable—perhaps to mitigate risks of over- or under-deterrence, or to reduce the time and expense of rule-application, or to serve other reasonable systemic goals. For example, while most U.S. jurisdictions authorize sentencing officials to impose any sentence for voluntary manslaughter between two broadly spaced poles (say, 2-20 years), California authorizes only 3 possible sentences: 3, 6, or 11 years. CAL. PENAL CODE § 193(a) (West 2010). Closer to home, law students might contrast the 32-interval grading system used at the University of Chicago (any numerical score from 155-186, inclusive) with the 4-interval system employed at Yale (Honors/Pass/Low Pass/Fail).

[37] Of course, things are not quite so simple, and some dynamics might cut in just the opposite direction. Most notably, insofar as early scores might have especially pronounced effects on strategy and psychology, one might predict that, holding closeness of contest constant, points scored earlier in contests are likely to have a greater impact on outcome than points scored later. The effect of early scores on game progression is ultimately an empirical question and likely to vary substantially across different sports. Nonetheless, I suspect that the effects of early scores are less regular then this surmise suggests. Sometimes the team that goes down early will become demoralized, other times it will become refocused; sometimes the team that goes up will play with greater confidence, other times it will become complacent or sloppy. But if you're losing, you simply can't prevail unless you score more than your opponent in the time remaining. And the less time that remains the harder that it is to do.

A second qualification to the claim in text is best presented with an illustration. In a high-scoring matchup between two football teams with explosive offenses and little defense, it may often appear that whichever team gets the ball last will win. In such a contest, and within a certain hard-to-define temporal range, a go-ahead score by Team A might maximize its probability of victory if the score comes with more, rather than less, time remaining. Team A's best bet, of course, is to leave Team B with insufficient time to score. But failing that, Team A is better

way to think about the point is this: the less time remaining in the contest, the greater the impact of each unit change in score on "closeness of contest," and thus the greater the expected effect on outcome of each unit change in probability of a score.[38] "Crunch time" just is that period when everything matters more.

It is true that $\phi_{In\text{-}A(B)}$ and $\phi_{Pn\text{-}A(A)}$ both increase in crunch time, and that the ratio between the two probably remains constant. But the absolute magnitude of the difference between the two increases. And because $\phi_{Pn\text{-}A(A)}$ may substantially exceed $\phi_{In\text{-}A(B)}$, the change in the outcome-affecting difference between allowing the infraction to go unrectified and imposing a penalty might well justify heeding the call to "let 'em play"—certainly in cases of NHNP, and very possibly in cases of LHNP too.

F. An objection: risk damage and the lost chance doctrine in tort law.

The argument for temporal variance in the enforcement of fouls depends on the following four propositions, among others:

---

off leaving its opponent with too much time to exhaust, thus increasing the likelihood that if Team B does score, Team A will be left with a final possession of its own. I am fairly confident that the general phenomenon that this example instantiates is fairly unusual, but cannot examine its contours or incidence here. In the meantime, we might dub the phenomenon "The Upper Hand Caveat" because it depends upon the nonlinear structure of alternating opportunities found in the familiar method for determining which of two captains picks first in selecting teams for sandlot baseball. As one authority explains the ritual: "One puts a hand around the bat near the fat end, then the other puts a hand around the bat just above his hand. This goes on, hand over hand, until the bottom of the bat is reached and there is no room for another hand. The last hand on the bat wins the contest (although the loser does have the chance to delicately grasp with his fingertips whatever little wood is left and twist it around his head, winning if he can hold on to the bat while doing this three times)." ROBERT HENDRICKSON, THE FACTS ON FILE ENCYCLOPEDIA OF WORD AND PHRASE ORIGINS (1997).

[38] Peter King recognized that events gain in significance in inverse proportion to time remaining when analyzing Bill Belichick's much-discussed decision to go for it on fourth and 2 from his own 28 with 2:08 remaining, up 34-28 against Indianapolis in week 10 of the 2009 season. The Patriots didn't get the first down, the Colts took over on downs and scored the winning TD with 13 seconds remaining. King is critical of the decision, urging that New England should have punted instead. His criticism was misguided. (*See*, for a powerful defense of Belichick's decision, Frank Frigo, The Anatomy of a (Fourth-Down) Decision, http://fifthdown.blogs.nytimes.com/2009/11/25/the-anatomy-of-a-fourth-down-decision/?pagemode=print (Nov. 25, 2009).) But in the course of his analysis, he makes an astute observation. Acknowledging that Belichick had made a similar call earlier in the season that had worked out well, King notes a critical difference: "Against Atlanta in Week 3, there was a play something like this. New England had fourth-and-one at its 24 late in the third quarter, up 16-10. Sammy Morris ran for two yards, first down, and the Patriots went on to kick a field goal on the drive. But that was one yard, not two, and even if it had failed and the Falcons got the ball and scored, *the Patriots would have had an entire quarter to rectify things*." (my italics). Peter King, *No Matter Which Way You Dissect It, Bill Belichick Made the Wrong Call*, SI.COM, Nov. 16, 2009, http://sportsillustrated.cnn.com/2009/writers/peter_king/11/15/mmqb/index.html?eref=sihp

1. at t1, before an infraction, competitor B "has" or "faces" some probability, P1, of victory;

2. at t2, as a result of an infraction by competitor A, B faces probability P2 of victory;

3. P1-P2 = R, R> 0; and

4. the reduction in probability of victory, R, is a compensable injury or harm to B.

This argument is analogous to the argument in support of the "lost chance" doctrine in tort law, which provides that negligent conduct (action or omission) by some party that increases the probability that some person will experience some uncontroversial injury (usually, death, bodily injury, sickness, etc.) is *itself* a compensable injury. The argument for recovery runs like this:

1. at t1, individual B "has" or "faces" some probability, P1, of some specified bodily injury;

2. at t2, as a result of negligent conduct by A, B faces some other probability, P2, of that same bodily injury;

3. P2-P1 = R, R> 0; and

4. the increase in probability of injury, R, is an injury or harm to B.

Although the lost chance doctrine is recognized in the United States and Canada and has been supported by many tort scholars, Britain and several other scholars—Stephen Perry most influentially—reject it.[39] If Perry's arguments against lost chance in tort law are correct, and if the sports case is not distinguishable, then my argument fails. So ultimately I must establish either that Perry's argument is not entirely correct or that it is distinguishable, or both. Given the complexity of Perry's arguments, their somewhat changing shape, the difficulty of some of the underlying issues, and the length and exploratory character of this essay, I cannot attempt a

[39] The leading American and English cases are, respectively, Herskovits v. Group Health Coop., 664 P.2d 474 (Wash. 1983), and Hotson v. East Berkshire Area Health Authority, [1987] 2 W.L.R. 287, *rev'd* [1987] A.C. 750. Perry's articles on the subject include *Risk, Harm, Interests, and Rights, in* RISK: PHILOSOPHICAL PERSPECTIVES (Tim Lewens ed. 2007); *Harm, History, and Counterfactuals*, 40 SAN DIEGO. L. REV. 1283 (2003) [hereinafter Perry, *Counterfactuals*]; *Risk, Harm, and Responsibility, in* PHILOSOPHICAL FOUNDATIONS OF TORT LAW 321 (David G. Owen ed. 1995); and *Protected Interests and Undertakings in the Law of Negligence*, 42 U. TORONTO L.J. 247 (1992). *Compare also, e.g.,* Matthew D. Adler, *Risk, Death and Harm: The Normative Foundations of Risk Regulation*, 87 MINN. L. REV. 1293 (2003) (agreeing with Perry that risk damage is not a welfare setback) *and* Claire Finkelstein, *Is Risk a Harm?*, 151 U. PA. L. REV. 963 (2003) (arguing otherwise).

comprehensive evaluation here. This is just a sketch of the issues and some possible routes forward; it is inescapably preliminary and telegraphic.

The tort literature contains various arguments against recovery for lost chances, or what are better termed augmented probabilities of harm. All start from the idea that if augmented probability of harm is a harm in its own right—that is, a harm apart from such uncontroversial harms as the fear and anxiety the plaintiff might experience in contemplation of future harm, the costs the plaintiff might incur for medical monitoring, the life opportunities the plaintiff might forgo to reduce the danger of realizing the augmented risk, etc.—it must be an increase in objective probability of outcome harm, not an increase in probabilities variously described as subjective, epistemic, or Bayesian. From that point of departure, at least three distinct arguments can be discerned.

The first argument denies that there exist objective probabilities with respect to singular events or propositions at least with respect to events governed by deterministic causal processes. The dominant account of objective probability, frequentism, defines the probability of an outcome as the limit of its relative frequency in a long series of events or, put otherwise, as the ratio of the times that outcome occurs to the times it might have occurred. Because any individual event falls within indefinitely many reference classes, objective probabilities cannot uniquely extend from the reference classes to the individual event itself. As the leading theorist of frequentism, Richard von Mises, announced eighty years ago, an objective probability for singular events is "utter nonsense." Let's call this the "no objective singular probabilities" thesis.

The second argument may grant (at least arguendo) that objective probabilities can apply to singular propositions, but maintains: (a) that, given determinism, an individual's probability of suffering some harm in the future is always 0 or 1; (b) that an event cannot be harmful if it does not change the pre-existing objective probability of harm from 0 to 1; and (c) that, when some event does change the pre-event probability of harm from 0 to 1, that event causes actual unproblematic harm, leaving "[no] room for the idea that risk constitutes a separate form of harm."[40] Call this the "merger argument": an event augments the objective

[40] Perry, *Risk, Harm, Interests, and Rights, in* RISK: PHILOSOPHICAL PERSPECTIVES, *supra* note 39, at 196.

probability of outcome harm only when it causes that harm (by changing the probability of outcome harm from 0 to 1). But in that case the risk damage merges into the outcome harm itself.

The third argument grants that objective probabilities can apply to singular events and that such probabilities can lie between 0 and 1, and can change, but denies that any such increase in an objective probability of unproblematic harm (i.e., of a setback to "core interests") is itself a compensable harm (setback to core interests) for purposes of tort law. Call this the "not a welfare harm" argument.[41]

This last argument is no threat to my analysis for temporal variance for it depends entirely upon claims about human welfare or well-being and about the proper function of tort law. There is no inconsistency in maintaining that an augmented objective probability of physical injury or illness is not itself the sort of setback to interests that tort law should treat as a compensable injury while allowing that a diminished objective probability of victory is a setback to interests that competitive sports may be concerned to compensate. So insofar as Perry and others rely upon the "not a welfare harm" argument, the argument against the lost chance doctrine and my argument for temporal variance are easily distinguished.

The merger argument might prove harder to distinguish, but it is a bad argument because (on deterministic assumptions) the pre-event objective probability of outcome harm already takes fully into account the subsequent occurrence of the event in question (like a negligent medical misdiagnosis), thus making the pre-event and post-event probabilities of outcome harm the same. Put another way, if determinism is true, then any event that might seem to change an objective probability from 0 to 1 was itself determined and thus fully accounted for in the pre-event objective probability of post-event harm.

That leaves only the first argument as an obstacle to my argument. Unfortunately, that is probably the dominant philosophical argument against the lost chance doctrine. I believe I am left with two avenues of escape.

First, although the exposition to this point suggests an objective construal of singular probabilities, the analysis does not strictly rely upon it, but can make do with epistemic

[41] *See, e.g.*, *id.* at 201–04; Perry, *Counterfactuals*, *supra* note 39, at 1304–08; Adler, *supra* note 39.

probabilities. Consider the discussion in Section II.E that assumed that a team had some objective probability of victory less than 1 and greater than 0. If that is not so—if the objective probability was unchanging at either 0 or 1—it would still be true that the epistemic probabilities range between 0 and 1. Accounts of epistemic probabilities vary regarding the extent to which they would impose constraints on actual subjective probability estimates. On a radically subjectivist version, the epistemic probability that the Lakers will beat the Celtics might be .5 for you, .58 for a professional gambler, and .25 for a shamrock-eyed Celtics fan, and that is that. Most versions of epistemic probability, however, incorporate standards of evidence and reasonable inference that provide resources for critically assessing actual subjective probabilities as good or bad, or, at a minimum, better or worse. If there are intersubjectively valid epistemic probabilities, and if infractions and penalties change those probabilities in the ways Section II.E discusses, that is all that the argument of this section requires.

The second avenue of escape is more ambitious and therefore potentially more interesting. It is to challenge the categorical denial of objective singular probabilities. I confess that, with others,[42] I am influenced by pre-theoretical intuitions that I find hard to shake: the ubiquity, utility and apparent meaningfulness of objective singular probability (OSP) judgments are too great to be easily dismissed. But we can say a little more to sustain hope that the search for OSP is not a fool's errand. First, some theorists who have rejected OSP might have been too quick to reduce objective probabilities to relative frequency accounts of probability, paying insufficient attention to the possibility that alternative accounts of objective probability might be workable. In particular, they have overlooked "propensity" accounts of objective probability, a term initially coined by Karl Popper for his own theory but now routinely applied to the family of objective, non-frequency accounts.

Now, it is true that propensity accounts are disfavored by philosophers of probability. But—and this is an additional reason for hope—the reasons for the rejection may be more dependent on those philosophers' particular scientific aspirations than legal commentators who credit that rejection might appreciate. Consider in this vein Donald Gillies's observation

[42] *See, e.g.,* Finkelstein, *supra* note 39, at 997–98.

that, "while there is nothing wrong with developing a metaphysical theory of propensities," his "own aim is to develop a . . . theory of probability which can be used to provide an interpretation of the probabilities which appear in such natural sciences as physics and biology."[43] To satisfy that ambition, he explained, "probability assignments should be testable by empirical data, and this makes it desirable that they should be associated with repeatable conditions"—a desideratum that standard propensity theories cannot satisfy.[44] Gillies's concern might justify the rejection of propensity-based OSP by science, but it does not license the conclusion that OSPs are false.

III. Two Kinds of Faults, Many Kinds of Rules

At first blush, we might suppose that the analysis of Part II applies, mutatis mutandis, to foot faults in tennis and therefore that McEnroe's position is vindicated: tennis officials should call foot faults less strictly at crunch time. The penalty for foot faults should not have been imposed on Serena Williams even assuming that her foot did touch or slightly cross the baseline. But such a conclusion would be premature. It could be that foot faults in tennis differ from fouls and similar infractions in basketball, football and comparable sports[45] in ways that

---

[43] Donald Gillies, Philosophical Theories of Probability 128 (2000).

[44] *Id.*

[45] Hockey fans will have noticed that, although I have now mentioned the sport a small handful of times, I haven't weighed in on the much-debated questions of whether power plays should be awarded less liberally in overtime than during regulation or in the playoffs than during the regular season. There's a good reason for that: I don't watch a lot of hockey. So I'm offering my four cents in a footnote in the hope that, if what I say here is unusually daft, fewer readers will be aware of it.

First, it seems to me that the analysis provided so far does offer support for the view that refs should swallow their whistles in overtime. Second, the analysis probably also lends credence to the thought that the game should be called less tightly during the playoffs than during the regular season. We have thus far treated the relevant unit as the individual game: we want the outcomes of individual games to reflect the competitors' relative success in realizing the sport's core athletic excellences and in overcoming the sport's central challenges. But we could as well take the unit of competition to be the season. Insofar as teams are competing for season success—which would mean, in the NHL, the chance to compete for, and ultimately win, the Stanley Cup—then whatever argument might support temporal variance for overtime relative to regulation would likely translate fairly straightforwardly to support temporal variance for the playoffs relative to the regular season.

Third, and on the other hand, power plays in hockey are arguably different from all of the penalties we have discussed so far. We might think of free throws in basketball and yardage mark-offs in football as necessary evils. There is nothing particularly exciting or valuable about either; the leagues would do away with them both but for the need to compensate victims of infractions for the competitive harm that has been done them and to deter future such infractions. At least some people, however, view power plays differently. Indeed, some have speculated that, when, in 2006, NHL Commissioner Gary Bettmann instructed refs not to call the playoffs more

make a difference. This Part advances two arguments: first (in Sections A through C), that foot faults *do* differ in a way that matters; and second (in Sections D through F) that temporal variance in their enforcement can nonetheless be defended on alternate grounds.

A. From the hardwood to the diamond.

I claimed in the Introduction that the case for temporal variance is made difficult by clear cases for invariance. My example was strikes in baseball. Nobody, I supposed, would think that umpires should cut pitchers a little more slack toward the end of a close game. But was I too quick? The famed Harvard zoologist-cum-Yankees fan Stephen Jay Gould would have us believe so. Here's his account of Don Larsen's perfect game in the 1956 World Series—to this day, the only perfect game or no-hitter in postseason play—penned on the death of the home plate umpire that day, Babe Pinelli:

> Babe Pinelli was the umpire in baseball's unique episode of perfection when it mattered most. October 8, 1956. A perfect game in the World Series—and, coincidentally, Pinelli's last official game as arbiter. What a consummate swan song. Twenty-seven Dodgers up; twenty-seven Bums down. . . . [T]he agent was a competent, but otherwise undistinguished pitcher, Don Larsen.
>
> The dramatic end was all Pinelli's, and controversial ever since. Dale Mitchell, pinch hitting for Sal Maglio, was the twenty-seventh batter. With a count of 1 and 2, Larsen delivered one high and outside—close, but surely not, by its technical definition, a strike. Mitchell let the pitch go by, but Pinelli didn't hesitate. Up went the right arm for called strike three. Out went Yogi Berra from behind the plate, nearly tackling Larsen in a frontal jump of joy. "Outside by a foot," groused Mitchell later. He exaggerated—for it was outside by only a few inches—but he was right. Babe Pinelli, however, was more right. A batter may not take a close pitch with so much on the line. Context matters. Truth is a circumstance, not a spot.
>
> . . .
>
> Truth is inflexible. Truth is inviolable. By long and recognized custom, by any concept of justice, Dale Mitchell had to swing at anything close. It was a

---

loosely, he was motivated in part by the view that power plays are positive goods, not just necessary evils, because they increase scoring and scoring is always exciting.

Fourth and finally, it appears to me that much of the debate about calling penalties in the playoffs is not principally about temporal variance. I think—and am open to being corrected—that many of the folks who urge that the refs should "let them play" in the playoffs would prefer that they "let them play" during the regular season too. These fans simply prefer a more physical style of hockey. They focus their attention on the playoffs not so much (or not only) because they think the playoffs are meaningfully different but in conformity (conscious or not) with the maxim that one ought to pick one's battles.

> strike—a strike high and outside. Babe Pinelli, umpiring his last game, ended with his finest, his most perceptive, his most truthful moment. Babe Pinelli, arbiter of history, walked into the locker room and cried.[46]

Is Gould right that Pinelli should have called that non-strike a strike? Is this a case for temporal variance? I cannot do justice to Gould's remarkable essay in this space, so two brief points. First, whereas the case for temporal variance I have sketched depends entirely on an assessment of the costs and benefits to the competitors, Gould rightly draws attention, in addition, to the impact that calls and no-calls can have on others—fans at the time and even, we might say, posterity. Second, and nonetheless, if informal polling of students and colleagues provides reliable guidance, Gould is in a very small minority. Most people think that strikes and balls are not appropriate candidates for temporal variance even here, in what we might reasonably suppose is the best-case scenario for it. If Mitchell "had to swing" at a close 2-1 pitch—if a batter "may not" take a close pitch in that setting—that's because it's unwise to risk a mistake about the pitch's location, either by himself or by the umpire. The command is prudential not normative. Accordingly, that Mitchell should have been aware that the umpire might *erroneously* call a near-miss a strike doesn't mean that the umpire should have done so purposefully. Or, to put the thought in a nutshell: had Dale Mitchell been Ted Williams, custom and justice would not have demanded that he swing at a pitch he knew to be outside.[47]

If this is so, two tasks remain. First, we must examine whether the analysis developed in Section 4 that lends support for temporal variance in the enforcement of infractions like fouls in basketball can explain why slack is not appropriate in the calling of balls and strikes. If it can't, then we have reason to worry that we have someplace gone astray. But if it can, then we are poised to undertake the second and final task: to determine whether the analysis already

---

[46] Stephen Jay Gould, Op-Ed., *The Strike That was Low and Outside,* N.Y. TIMES, Nov. 10, 1984, at 23. Gould's piece was mistitled. Video of the play makes clear, as the body of the essay suggests, that if the pitch wasn't the right height, it was too high, not too low.

[47] The Splendid Splinter's extraordinary vision and command of the strike zone—and the respect it earned him from the men in blue—is often illustrated with the story of the young catcher who complained that a close pitch was miscalled a ball. "Son," the ump is said to have replied, "when the pitch is a strike, Mr. Williams will let you know." *See, e.g.,* ZACK HAMPLE, WATCHING BASEBALL SMARTER: A PROFESSIONAL FAN'S GUIDE FOR BEGINNERS, SEMI-EXPERTS, AND DEEPLY SERIOUS GEEKS 122 (2007). In point of fact, many authorities attribute the quote to Hall of Fame umpire Bill Klem—and about Rogers Hornsby, not Williams. But that's nitpicking: Williams had a pretty fair eye, and the umps knew it.

advanced, refined or supplemented as may prove necessary, applies to foot faults. Speaking loosely, the question will be whether foot faults in tennis are more like fouls in basketball or like balls in baseball.

B. Two kinds of rules

The analysis for temporal variance in the enforcement of penalties for fouls relied upon "no harm, no penalty" reasoning. We said that there are times when it might better serve the objectives of competitive sports to refrain from enforcing a penalty despite the occurrence of an infraction. That's because the competitive costs of an infraction and of the sanction or penalty that it begets are both temporally variant and the latter can become, at game's end, very much greater than the former. Yet assessing the competitive costs of these two things—the infraction and the sanction—seems impossible in the case of balls and strikes. It's impossible because the denomination of a pitch as a "ball" is not properly conceptualized as the penalty for an infraction; the concepts of infraction and penalty just don't apply here.

That not all undesired consequences that attach to nonconformity with the dictates of a rule are sanctions imposed for infractions was, of course, a central claim upon which Hart relied when critiquing the Austinian command theory of law. Most of the rules of the criminal law impose duties and threaten sanctions for their violation. But other legal rules, like those specifying the conditions for valid wills or contracts, are of a different sort. These, Hart proposed, are "power-conferring rules"—rules that (somewhat simplified) provide that "if you wish to do this, this is the way to do it."[48] In the case of rules that impose a duty, he explained, "we can distinguish clearly the rule prohibiting certain behaviour from the provision for penalties to be exacted if the rule is broken, and suppose the first to exist without the latter. We can, in a sense, subtract the sanction and still leave an intelligible standard of behaviour which it is designed to maintain."[49] But the distinction between the rule and the sanction is not intelligible in the case of power-conferring rules. It makes sense to say "do not kill" even when we leave off the part about what happens if you do. In contrast, we know we're leaving

[48] HART, *supra* note 7, at 28.
[49] Id. at 34.

something critical out of the picture if we say "get two witnesses" but don't explain that the will will be invalid otherwise.

The Hartian analysis of power-conferring rules helps to explain why balls and strikes in baseball feel very different from the infractions we have considered in basketball and football. In the case of the latter, we can sensibly ask both whether some type of contact ought to be proscribed (thus denominated as a "foul"), and, *in addition*, whether, if so, the penalty attached to commission of the foul—two free throws, say, or ten yards—is too great (or too small). But every pitch is either a ball or a strike. The logical consequence of its being outside the strike zone is that it is a ball. While we can sensibly ask whether the strike zone is too small (or too large), or whether the number of balls that constitutes a walk is too great (or too small),[50] or whether *any* number of balls should result in the award of a base,[51] it seems nonsense to ask whether a pitch's being a ball is too high a price for its having narrowly missed the strike zone: that the pitch was a ball is just what it *means* for its not having been a strike.

In short, we might provisionally endorse the following conclusion: Contra Gould, balls and strikes are not proper candidates for temporal variance because (1) temporal variance depends upon the widening of a gap between the competitive cost of an infraction and the competitive cost of the penalty it incurs, but (2) there is no such gap between nonconformity with a power-conferring rule and the consequences that attach, and (3) the rules governing balls and strikes are power-conferring rules (or something of a sufficiently close type).[52]

---

[50] In fact, in the early days of baseball, more balls were required to make out a walk. Five successive rule changes adopted from 1879 to 1889 dropped the number from nine to the current four. GLEN WAGGONER ET AL., SPITTERS, BEANBALLS, AND THE INCREDIBLE SHRINKING STRIKE ZONE: THE STORIES BEHIND THE RULES OF BASEBALL 114 (rev. ed. 2000); DAVID NEMEC, THE OFFICIAL RULES OF BASEBALL ILLUSTRATED 22 (2006).

[51] *See id.* (noting that early baseball had balls before it had bases on balls).

[52] The parenthetical is intended to signal that I am not wholly committed to this particular typology of rules. Perhaps, for example, the roughly analogous distinction between regulative and constitutive rules, most prominently associated with John Searle, might provide the more useful analytical framework. And I am open to other possibilities as well. (Roughly, "regulative rules regulate antecedently or independently existing forms of behaviour" whereas constitutive rules "do not merely regulate, they create or define new forms of behaviour." JOHN SEARLE, SPEECH ACTS 33-34 (1969). And they create new forms of behavior—what Searle sometimes calls "institutional facts"—by assuming the form "*X* counts as *Y* in context *C*." *Id.* at 52. Thus, for example, "moving the king two squares toward a rook, and moving that rook to the square over which the king has crossed counts as castling in chess." *See also, e.g.,* JOHN SEARLE, THE CONSTRUCTION OF SOCIAL REALITY (1995); RAIMO TUOMELA, THE PHILOSOPHY OF SOCIAL PRACTICES: A COLLECTIVE ACCEPTANCE VIEW (2002). For criticisms of the regulative/constitutive distinction, see, e.g., Frank A. Hindriks, *Constitutive Rules, Language, and Ontology,* 71 ERKENNTNIS (2009); Christopher Cherry, *Regulative Rules and Constitutive Rules,* 23 PHIL. Q. 301 (1973).)

C. Two kinds of faults

If the analysis of balls and strikes is correct, then the question whether temporal variance would have been appropriate in the Williams-Clijsters match might depend on how we conceptualize the rule governing foot faults—as duty-imposing or power-conferring (subject to the qualification in the preceding footnote). If the rules command that the server not step on or past the baseline, on pain of the serve being declared a nullity, then the reasoning that supports temporal variance in the calling of fouls would seem to be available, providing a possible basis for temporal variance in the calling of foot faults too. If, instead, the rules confer upon the server two opportunities (lets aside) to put the ball in play, and specify that the only valid way to put the ball in play includes that one not step on or past the baseline, then the analysis proffered in Part II would not apply, and we would have, thus far, no basis for temporal variance.

Which is the better conceptualization of the relevant tennis rules is, I think, more open to argument than one might suppose. There is no simple test that straightforwardly or uncontroversially resolves the question.[53] So let's start not with foot faults but with the more common type of service fault—the fault produced by the failure to strike the served ball into the diagonally opposite service court. Call this a "zone fault."

It seems as clear as these things can be that the rules of tennis are not rightly understood to impose upon a server the duty to strike the ball into the service court. To be sure, servers must be under a duty to *try* to put the ball lawfully or validly into play, for part of what it is to play a competitive game is to assume an obligation to compete. That is why if your

[53] The text of the relevant rules is always a good place to start. Rule 18, recall, is written in deontic terms, providing inter alia, that "during the service motion, the server shall not . . . touch the baseline or the court with either foot." Furthermore, Rule 19(a)– by providing that "the service is a fault if the server breaks rules 16, 17, or 18"—reinforces the idea that the server has a duty not to step on the line, for, strictly speaking, one does not "break" a power-conferring rule by failing to comply with its dictates. However, we should not place too much weight on constructions that are unlikely to have received much conscious thought by the drafters. After all, rule 17—a second rule that rule 19 contemplates being broken—employs similar deontic language in providing that "the serve shall pass over the net and hit the service court diagonally opposite." Yet no tennis player would think himself under a duty to hit the diagonally opposite service court. To the contrary, all players and fans understand that the server is empowered to put the ball into play by serving into the correct space. Because the plain function of rule 17 is to confer a power, not to impose a duty, we must characterize it as a power-conferring rule, the somewhat infelicitous language notwithstanding. We ought not then invest much faith in the ordinary meaning of the text of rule 18 either.

opponent doesn't even try to do those things necessary to score points, you will rightly object notwithstanding that her failure to attempt to satisfy the power-conferring rules all but assures you of victory. However a sincere but unsuccessful effort to strike the ball into the specified zone is not a violation of an obligation. It is an invalid serve—a "fault"—for the same reason that a will signed by only one witness is legally invalid and that an unswung-on pitch that misses the strike zone is a ball: the actor has failed to do what is specified to perfect a power. If, contrary to Gould, Babe Pinelli should not have punched Dale Mitchell out on Larsen's 1-2 pitch, then a line judge should call any serve that's long or wide a service fault regardless of game context.

Although the issue is more debatable, on balance it seems sensible to characterize the rules defining foot faults as power-conferring as well. In order to successfully or "validly" put the ball into play, thus giving oneself an opportunity to win the point, the server must do several things: (1) start behind the baseline, (2) strike the ball before stepping on or over the baseline, and (3) by striking the ball, cause it to land in the service court diagonally opposite. We might say that these are three components of the rule that defines a valid serve. A failure on any of these three grounds is just a failure to perfect the power conferred upon the server; none is a violation or an infraction.

That seems right.[54] But here's the puzzling thing. If foot faults are also governed by power-conferring rules, and if temporal variance could be defended only on the analysis developed to this point, then we should expect foot faults to be immune from temporal variance just as surely as are zone faults. But widespread intuitions are more equivocal. I have not run across anybody who is tempted by temporal variance for zone faults. If, facing match point, the server hits a second service wide by a smidgen, well them's the breaks and that's the match. And yet we have already seen that some—John McEnroe, for example—believe that foot faults should be enforced with temporal variance. Just as revealingly, many more feel that the temporal variance of foot faults is, at the least, more plausible, less obviously mistaken. The fact that even those who resist temporal variance for foot faults do not feel about foot faults quite as they do about zone faults—the fact that many of them at least feel the *tug* of

[54] It is also, as will become apparent, by far the more interesting assumption to indulge.

temporal variance—requires explanation even if we end up concluding that, all things considered, foot faults should be enforced invariantly.[55] That fact is inexplicable if the argument for temporal variance depends upon the widening of a gap between infraction and penalty and if faults aren't penalties for infractions.

I favor our taking widespread intuitions seriously. Doing so invites us to consider whether the analysis supplied thus far furnishes the *only* sound basis for temporal variance. Perhaps it doesn't. Perhaps temporal variance for some power-conferring or constitutive rules might be warranted on other (possibly related) grounds.

D. Athletic virtues revisited

Recall what we said earlier: We want the outcome of athletic contests to depend (insofar as possible) upon the competitors' relative excellence in executing the particular athletic virtues that the sport is centrally designed to showcase, develop and reward.[56] Call this "the competitive desideratum." It was not a stray or peripheral observation, but rather a linchpin in the case for temporal variance in the enforcement of penalties for incidental infractions in basketball and football. We only ask whether imposition of a penalty unduly affects the competition's outcome when the rule in question does not implicate the core athletic virtues and excellences of the sport in question. *If the rule does implicate the skills that the sport is focally designed to test, then to enforce the rule tends to satisfy rather than frustrate the competitive desideratum.* Thus we really have two reasons against slack in the calling of balls and strikes. First, as we have seen, calling a non-strike a "ball" is not a penalty

[55] I earlier noted the possibility that temporal variance is warranted by aesthetic considerations: when energy, intensity and drama have risen, we don't want delays and we want things to be resolved by excellence, not errors or miscues. While conceding that there is something to this analysis, I contended that it's not the whole story. Contemplation of foot faults and zone faults bolsters that suggestion. We prefer the match to end with a winner than with an unforced error (including a service fault), but we are *much* less tempted to temporal variance for zone faults and other unforced errors than for foot faults.

[56] I believe that the Supreme Court had something much like this in mind when considering whether the Americans with Disabilities Act required the PGA Tour to accommodate a pro golfer's physical disability by permitting him to ride a golf cart during competition. *PGA Tour, Inc. v. Martin*, 532 U.S. 661 (2001). En route toward concluding in the affirmative, the majority determined, in effect, that the central athletic challenge the PGA Tour presented was (to a first approximation) the ability to hole a ball by means of striking it with a club, in the fewest number of strokes, while battling fatigue. Justice Scalia, in dissent, charged the majority with presuming to opine on the nature of "Platonic golf." He was mistaken. The majority was trying to determine what was central or peripheral to the sport of golf *as constituted by actual PGA rules and practices*. That's an interpretive task, and any answer is almost certain to be reasonably contestable. But it is not an inquiry into Platonic essences.

for a rule infraction, but just part of the power-conferring rules of the sport. Second, the ability to throw a strike just is one of the central athletic challenges in baseball. (Having the eye to lay off a non-strike is a secondary excellence.) So calling strikes as "strikes" and no-strikes as "balls" promotes our goal that the outcome depend upon the teams' relative excellence in meeting athletic challenges that baseball centrally presents.

What about the rules governing serves in tennis? What are the athletic excellences or challenges involved here? To a first approximation, one of the core challenges is to strike the ball with power and accuracy into a predetermined space. This seems on the right track, though it's clearly not all the way home. Suppose the server were to stand at the net before striking the ball. Serving the ball into the service court from there plainly does not satisfy or conform to the athletic challenge that serving in tennis is meant to present. So a refinement is necessary. Perhaps this: the challenge is to strike the ball *into a precisely defined space from a precisely defined distance*.

I'm going to suggest that this is not in fact the best rendition of the athletic challenge that the service rules are meant to embody, and that the challenge is better formulated as the ability to serve the ball *into a precisely defined space from a generally defined distance*. That is, notwithstanding that the formal rules appear to specify both the starting point and the landing space with precision, the "real" or underlying athletic challenge that the rules are designed to codify or facilitate involves a precise target but an imprecise or general launching site. I am tempted to describe the challenge this way: "get the ball *in here* from *around there*." That is surely putting things too loosely, but it conveys my basic claim that precision in the placement of the served ball is of far greater concern to the sport than is precision in the placement of the server's body.

I'll try to make this asymmetry plausible in the remainder of this section. The final two sections will explain why this claim, if true, might support temporal variance for foot faults but not for zone faults.

Let's start with zone faults. Why is the placement of the ball regulated precisely? Why isn't the athletic challenge to get the ball close enough to the service court? Tennis requires that the ball go into the service court because that's the athletic challenge that serving in tennis

is designed to (if you'll pardon the pun) serve up. It is how tennis instantiates one of the most commonly tested skills across all of sports: target-hitting. And horseshoes and curling notwithstanding, precision is generally part of the nature of targeting—for pitchers, field-goal kickers, basketball players, no less than for archers and riflemen. To be sure, there is rarely anything essential about the target dimensions. But while the target's contours may be arbitrary, the demand that the competitor hit the target, as it were, and not merely come close, is not arbitrary, for the rule is designed to test and reward that particular class of physical excellences involving accuracy and precision in limb-eye coordination.[57] The rules of tennis require that, for a serve to be valid, the ball must land within the defined service court because that is part of the nature of this particular athletic challenge.

If that's so, why doesn't the same reasoning apply at the front end too? Why isn't precision required with respect to the placement of the server's feet at the moment of striking the ball just as surely as it is required with respect to the location of the ball's landing spot? My claim is *not* that it *couldn't* be. That could be the best understanding of the athletic challenge. I mean only to argue that it needn't be and to suggest that it probably isn't.

My suggestion will appear more plausible if we start by assuming the opposite. Suppose the athletic challenge that underlies or motivates the rules involves precision at the front end as much as at the back end. What would precision here involve? To start, why should it involve the *feet* at all? If the challenge were to serve the ball into a specified space from a specified distance, why isn't the relevant distance the distance that the *ball* must travel? Why wouldn't the challenge be better understood to require that the racquet strike the ball behind the vertical plane defined by the baseline? Furthermore, even if the specific distance that should matter is the distance from net to feet, why should we care about the precise location of the

[57] Consider some spectacular display of ball-handling artistry in which a soccer player dribbles the ball from deep in his own end up the pitch, deking and dodging defenders all the way until, 30 yards from the goal, he blasts a frozen rope through a tiny window between two defenders that just catches the inner edge of the right goal post and ricochets across the goal mouth before skittering out of bounds. Fans of the sport who are not partisans of either team might well wish that the shot had scored, to reward the player for his dazzling skill. Yet nobody would argue that a goal should be counted. While we might have made the goal a touch larger—indeed, while we might have good reasons for thinking that, all things considered, slightly enlarging the goal would improve the sport—we understand that the challenge is to put the ball into the goal, and not to get it "close enough."

feet at the moment the racquet strikes the ball and not be satisfied with specifying their location at the start of the service motion?

That the feet must remain behind the baseline would make sense were part of the challenge of the serve to strike the ball while keeping one's feet stationary. But that's not the case. While the written rules of tennis forbid the server from walking or running while serving, they expressly permit "slight movements of the feet." The rules-in-action, moreover, clearly allow much more than that. Given that the sport (sensibly) allows the server to move her feet during what is, after all, a powerful and explosive movement, why should it deem the *precise* location of the feet at the moment of impact of any importance? While some sports care very much about full bodily control—diving and gymnastics come immediately to mind—tennis is not such a sport. Tennis is a sport of speed and power, hand-eye coordination and overall athleticism. Indeed, the sport is so unconcerned with serving mechanics as even to permit underhand service. It seems much more in keeping with the type of athletic enterprise that is tennis to allow for natural service motions, not to micromanage the placement of the feet.[58]

In short, it seems more faithful to the type of sport that tennis is and to the nature of the service to understand the underlying athletic challenge as I have described it: to strike the ball into a precisely defined space while starting from a defined place and not traversing unreasonably past the line.

But here's the thing. The rules of tennis don't say that an attempted serve is a fault if the server steps "unreasonably far over the line" or anything of this sort. They say, as we have seen, that the serve is a fault if, during the service motion, either foot touches the court including the baseline. Why isn't that the end of it? We needn't speculate as to the nature of the athletic challenge in tennis as presently constituted, the objection goes, for the rules tell us exactly what the challenge is.

To address this complaint lodged on behalf of temporal invariance will require that we introduce one final distinction.

[58] I think that football is different in this regard. Full bodily control is part of the athletic excellence demanded of the wide receiver, which is what made Santonio Holmes's tip-toe touchdown reception in the 2009 Super Bowl such an impressive feat.

E. Two more kinds of rules

The distinction between duty-imposing and power-conferring rules is one of function. Another, more common, way to categorize rules—it might be clearer to speak now of "norms"—is on the basis of form. This is the division between rules and standards. True, the rule/standard distinction describes something closer to poles on a continuum than to binary possibilities, and any given complex norm can consist of more rule-like and more standard-like pieces cobbled together. Still, the basic difference between the two is fairly well settled: rules turn upon factual predicates that are sharper-edged, whereas standards require those who apply them to exercise evaluative judgment. Also well accepted are the considerations that recommend proceeding with one form or the other. To a first approximation: standards better reflect the genuine justifications that underlie the norm, while rules, because they are quicker and easier to apply, promote second-order considerations in cheap, predictable, stable, uniform, and non-corrupt decision-making. As Fred Schauer summarized: "the choice of rule-based decision-making ordinarily entails disabling wise and sensitive decision-makers from making the best decisions in order to disable incompetent or simply wicked decision-makers from making wrong decisions."[59]

Consider the speed limit. There's nothing magical about 65 m.p.h. or 55 m.p.h. or what have you. The real, true, or underlying norm is that people shouldn't drive dangerously fast—or, because a steel contraption weighing one to two tons is dangerous when driven at just about any rate of speed, they shouldn't drive *unduly* dangerously fast. But were we to announce that standard as the norm, we'd end up with a heterogeneity in rates of travel (the unusually timid driving too slow, the overly confident driving too fast) that might itself be unsafe. Furthermore, we'd impose additional decisional burdens on police and invite challenge to every enforcement action. Because we think we can identify a numerical speed limit that

[59] FREDERICK SCHAUER, PLAYING BY THE RULES: A PHILOSOPHICAL EXAMINATION OF RULE-BASED DECISION-MAKING IN LAW AND IN LIFE 153 (1991). *See also, e.g.,* Larry Alexander & Emily Sherwin, *The Deceptive Nature of Rules,* 142 U. PA. L. REV. 1191 (1994) (describing rules as lies).

approximates the true contextual line between safe and unsafe driving tolerably well, albeit imperfectly, we almost always proceed here by rule rather than by standard.[60]

F. True rules and rulified standards

We can now see that the fact that the rule governing foot faults is written in hard-edged terms—a foot fault is defined to exist if either of the server's feet even touches the baseline—is not inconsistent with my claims that the real norm that the rule is designed to implement might be a standard that prohibits servers from going "too far" over the line. Even if the true norm is a standard, it doesn't follow that the formal norm should assume the same shape. Quite the contrary. Because the factors that bear on reasonableness would be debatable in every case, considerations like predictability, certainty, and finality all forcefully favor implementing this norm by means of a rule rather than by means of a standard.

But the part of the service rules that require the ball to land within the service court is different. That, we agreed, is part of the underlying athletic challenge. The written criteria of valid service that govern the landing of the ball and the placement of the server's feet are, in both cases, rules rather than standards. But the former is a rule because it captures an aspect of the underlying athletic challenge that is *itself* sharp-edged and rule-like: get the ball *in* the pre-defined space. The latter is a rule because, although the aspect of the underlying athletic challenge that it captures is standard-like—start from behind the line and don't go unreasonably over it—we have good institutional reasons to codify it in bright-line terms. To coin terms, we might say that that portion of the power-conferring rule of tennis service that requires the serve to land in the service court is a "true rule," whereas that portion of the rule that requires the server not to step on the baseline is a "rulified standard."

Even assuming all this is so, what follows? To start, does it follow that line judges should enforce the rule governing faults as though a foot fault could occur only when the server steps unreasonably far over the line?

---

[60] We don't always do so, but the rare exceptions tend to bolster the point in text. In the late 1990s, Montana eschewed a rule-like speed limit in favor of a standard that mandated driving "at a rate of speed no greater than is reasonable and proper under the conditions existing at the point of operation . . . so as not to unduly or unreasonably endanger the life, limb, property, or other rights of a person entitled to the use of the street or highway." That proved to be a short-lived experiment.

Surely not. A rulified standard is a rule, not a standard, and enforcement authorities should generally apply it as such. To routinely pierce the rule and apply the underlying or animating standard would defeat the purposes served by having rulified it in the first place. That must be common ground. It need not, however, be the end of the story.

First, that we must not *routinely* pierce a rulified standard does not mean that we must *never* pierce it. Whether and under what circumstances to disregard the rule's form in favor of its underlying considerations is always at least askable with regard to rulified standards. Indeed, that is the most obvious upshot of the distinction between these two types of rule.[61]

Second, it is plausible to suppose that two additional requirements should be satisfied in order to go beneath the surface of a rulified standard: (1) that enforcing the rule as a rule would produce unusually high costs; and (2) that disregarding the rule's form on this occasion would incur low costs on the dimensions, such as predictability and the like, that justified its rulification in the first place.

These two additional conditions, it seems to me, are plausibly satisfied by foot faults in crunch time. The high cost of enforcing the rule as a rule are plain: doing so allows the foot fault to have an undue impact on the match outcome—that is, it thwarts what we have called the "competitive desideratum"—thereby detracting from the participants' satisfaction and the spectators' enjoyment. At the same time, the costs of piercing the rule are very low precisely because the fact of the supposed nonconformity with the rule is hidden from public view. And it's hidden from public view because the Hawk-Eye electronic system that determines whether a ball lands within the lines is not used to judge foot faults. From the perspective of optimal game design, that might be a good thing. In general, rule makers who want to preserve the rule-enforcers' option to sometimes apply the standard that animates a rulified standard should arrange things so that non-compliance with the rule isn't apparent. Transparency is not always a virtue.

---

[61] This is not to say, however, that true rules must always be adhered to. Claire Finkelstein has persuasively argued that exceptions to a rule are best conceived as reflections of purposes or principles external to the purposes, principles, or considerations that underlie the rule itself. Claire Oakes Finkelstein, *When the Rule Swallows the Exception, in* RULES AND REASONING: ESSAYS IN HONOUR OF FREDERICK SCHAUER (1999). On this view, true rules and rulified standard alike can be overridden by an exception. My claim is that, for rulified standards but not true rules, we can resist the rule's directive in a second way too—by disregarding its form in favor of the rule's own animating reasons.

Of course, even if an ideal system would have (non-expressly) authorized line judges to adjudicate crunch-time foot faults against the underlying standard of reasonableness and not in terms of the nominal rule, that does not fully determine whether Serena Williams's step on the line should have been called. It could be that it was unreasonable or unfair all things considered—if, for example, her transgression was substantial or repeated. My sense is that it was neither, but I make no strong claims about it. I claim more strongly that Williams's step on the line did not apparently put Clijsters at a competitive disadvantage: the ball landed squarely in the service court and was easily returnable. In sum, if I'm right that the foot fault rule is a rulified standard not a true rule, that would be a promising (though not conclusive) basis for supporting the McEnrovian intuition: the line judge should have cut Williams some slack.

CONCLUSION

My analysis of the puzzle of temporal variance in sports turns out to be two analyses, not one. Temporal variance in the enforcement of penalties for infractions can be explained and justified by the fact that the over-restitutionary impact of penalties is greater as time toward contest completion diminishes. Temporal variance in the enforcement of power-conferring or constitutive rules depends upon a distinction between two types of rules (as contrasted with standards)—true rules, and rulified standards—married to the propositions that context might warrant disregarding the latter in favor of its underlying standard, and that a "technical" foot fault might comply with that standard even though it runs afoul of the implementing rule.

These arguments are tentative and partial, only first steps toward a solution to the puzzle. But whether they ultimately justify the temporally variant enforcement of particular rules of particular sports, all things considered, will strike most jurisprudentially minded readers as of secondary importance. The arguments here have expressly drawn on practices and analyses from law and legal theory. Moreover, they offer significant promise of returning the favor—offering insights of value for, e.g., the lost chance doctrine in torts; the granting of equitable relief, near game's end, from rules governing municipal and corporate elections, or appellate litigation; the difference between genuine "jurisdictional rules" and mere claim-

processing rules; and possibly much else. Those promissory notes can't be cashed in this essay. But readers sensitive to the depth and complexity of the philosophical puzzles that arise on the fields of play have reason to suspect that sports will richly repay searching jurisprudential attention.