

REPLAY

Mitchell N. Berman

Abstract

This paper explores a question of superficial triviality: when sports use instant replay technology to review on-field calls, what standard of review should they employ? The conventional view is that on-field calls should be entrenched against reversal such that, if the reviewing official has any doubt about the correctness of the initial call, he must let it stand even if he thinks it very probably wrong. Indeed, in the wake of officiating debacles at last summer's FIFA World Cup, commentators proposed not only that soccer employ instant replay, but also that it follow the NFL in directing officials to overturn on-field calls only when "indisputable visual evidence" (IVE) reveals that call to be mistaken. This essay argues that common wisdom in favor of IVE overlooks important considerations against entrenchment and likely rests upon mistaken premises, and it offers several concrete proposals for reform.

A lengthy investigation into the optimal standard of review for instant replay in sports might seem frivolous. But it serves a deeper ambition. We are in the early years of sports' colonization by econometricians, as legal theorists remain watching from the sidelines. That is unfortunate. Formal organized sports are, in effect, legal systems, and legal theorists might find much both to teach and to learn by paying closer attention to competitive athletics. In short, legal theorists would benefit from a sustained engagement with what I have termed, in previous work, "the jurisprudence of sport."

As a case study in this nascent field, this essay reveals that the problem of appellate review in sports is surprisingly rich and complex—implicating profound questions concerning the relationships among desert, entitlement and justice; the difference between mistake and error; and the contours of loss aversion and omission bias, among other things. But it shows even more than that. The jurisprudence of sport maintains that sporting practices and norms can teach lessons for ordinary legal systems as surely as the other way around. Illustrating that claim, this essay draws from football replay practices an argument to reform the criminal trial system to accommodate two verdicts of acquittal, not one.

REPLAY

Mitchell N. Berman*

INTRODUCTION

By most accounts, the 2010 FIFA World Cup was a brilliant success for the host nation. Ever since soccer's world governing body had awarded the 2010 tournament to South Africa, making it the first African nation to host the world's largest sporting event, critics had worried that it would not be up to the task. As late as 2006, influential soccer figures had talked openly about withdrawing the tournament from South Africa over concerns that the stadiums would be incomplete or substandard and that crime and violence would run unchecked.¹ In the event, those fears went unrealized. Praising the quality of the facilities, the vigilance of the government, and the warmth of the people, FIFA President Sepp Blatter gave South Africa a grade of 9 out of 10, or "summa cum laude."²

If South Africans had reason to be proud of their accomplishments, fans across the globe gave the five-week tournament decidedly low marks. One common complaint focused on the vuvuzelas, a South African horn blown by fans that reaches an ear-splitting 127 decibels and whose drone has been compared to the mind-wracking buzzing of millions of angry bees. Even worse, though, was the officiating, which fans and journalists around the world derided as "an absolute joke."³

Fans of the American squad best remember referee Koman Coulibaly's never-explained disallowance of what likely would have been the Yanks' game-winning goal against Serbia off a direct kick by Landon Donovan. But the knockout round witnessed even bigger howlers. England, for example, was denied an equalizer against Germany when referee Jorge Larrionda failed to see

* Richard Dale Endowed Chair in Law, Professor of Philosophy (by courtesy), the University of Texas at Austin. Email: mberman@law.utexas.edu. [Acknowledgements to come]

¹ *State on Standby to Help Host 2010 World Cup*, THE AUSTRALIAN, Nov. 22, 2006, available at <http://www.theaustralian.com.au/news/state-on-standby-to-help-host-2010-world-cup/story-e6frg7s6-1111112566130>

² *2010 FIFA World Cup: South Africa Revels in Success*, TAIPEI TIMES, Jul.14, 2010, available at <http://www.taipetimes.com/News/sport/archives/2010/07/14/2003477885>.

³ Jack McCallum, *Refereeing at the World Cup is a joke, and it's hurting image of the game*, SPORTS ILLUSTRATED, June 28, 2010, available at http://sportsillustrated.cnn.com/2010/writers/jack_mccallum/06/28/world.cup/index.html.

that a shot by Frank Lampard that hit the underside of the crossbar landed a full two feet inside the goal. And Argentina took a 1-0 lead against Mexico on a goal by Carlos Tevez scored after Tevez had been at least a yard offside.

Gross officiating errors like these—following on the heels of a gaffe during qualifiers that allowed France to make it to South Africa on a blatant but uncalled Thierry Henry handball—provoked worldwide calls for soccer to initiate some form of replay. Public opinion on the question was so strong that even Blatter, who, only two years earlier, had refused even to consider proposals to employ technology to improve or review on-field officiating, felt compelled to change his tune. After having apologized to England and Mexican officials for the gross miscues in their matches, Blatter acknowledged that “we have to open again this file, definitely.” “Something,” he conceded, “has to be changed.”⁴

Remarkably, just a week before opening ceremonies in Johannesburg, Major League Baseball had been hit with an instant replay controversy of its own when umpire Jim Joyce robbed Detroit Tigers hurler Armando Galarraga of a perfect game with a mistaken call at first base with two outs in the ninth inning. The popular response was enormous, debate spilling over the sports pages and even hitting page one of the *New York Times*. Just as his World Cup counterpart would do later that same month, MLB Commissioner Bud Selig admitted that the sport would have to reconsider whether to expand use of instant replay beyond determining whether a ball hit over the fence was fair or foul.⁵

As technology advances, proposals to improve sports officiating by technological means—from using instant replay to review calls made by officials, to deploying electronic eyes that would displace initial human judgments—will only gain steam. Stakeholders of every variety—fans, players, owners, coaches, and broadcast partners—have already weighed in. But lawyers and law professors have not been particularly vocal or visible contributors to the debates.⁶ This essay aims to rectify that imbalance.

I have argued elsewhere that formal organized sports systems have much more in common with ordinary legal systems than is generally appreciated and, therefore, that legal theorists might find much both to learn and to teach by paying

⁴ Graham Dunbar, *Bad Calls Prompt FIFA to Consider New Technology*, ASSOCIATED PRESS, Jun. 29, 2010, available at http://www.salon.com/sports/feature/2010/06/29/world_cup_bad_calls_technology.

⁵ Jerry Crasnik, *Selig Doubts Replay Use Will be Expanded*, ESPN, Jun. 8, 2010, available at <http://sports.espn.go.com/mlb/news/story?id=5262731>. He also predicted that nothing would be changed.

⁶ But see *infra* note 26.

closer attention to the world of sports.⁷ Like legal systems, organized sports systems—by which I mean, to a first approximation, sports that feature written rules, rulemakers, and (at least at some levels of play) rule-enforcing officials—seek to facilitate and regulate behavior principally by means of general rules laid down in advance, rather than by individualized directives. That much is obvious. But congruence between the two domains run broader and deeper than is usually appreciated.

Like ordinary lawmakers, gamewrights confront virtually the entire panoply of problematics that traditionally engage legal theorists: when to regulate behavior “directly,” by means of formal written norms or “indirectly,” by refining and strengthening informal social norms; how best to navigate tradeoffs between rules and standards; when to delegate discretion to enforcement authorities, and how best to constrain it; how to respond to the problem of epistemic uncertainty; 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 much more. Given the great many similarities in the challenges that sports and law confront and the tools at their disposal, sports would richly repay more systematic attention from legal theorists and comparative lawyers. In recent years, econometricians have increasingly turned the powerful analytical tools of their discipline on sports.⁸ It is past time for jurists to follow suit.

The use of instant replay technology to review and possibly overturn calls made by officials on the field of play is one possible component of a system of appellate review of decisionmaking. It is therefore a natural subject of study for

⁷ See Mitchell N. Berman, “Let ‘em Play”: *A Study in the Jurisprudence of Sport*, 99 Geo. L.J. — (forthcoming 2011) (examining whether, and under what circumstances, rules of sports should be enforced less strictly at crunch time). This is not to claim that jurists have ignored sports (and games) entirely, but only that the attention paid to this domain—including by such giants as Hart and Dworkin—is occasional and ad hoc. See, e.g., HLA HART, *THE CONCEPT OF LAW* 89 (2d ed. 1994); RONALD DWORKIN, *LAW’S EMPIRE* 136-38 (1986). The best known precursors in the law reviews to serious jurisprudential or comparative engagement with both sports and law are more whimsical than probing. See, for the best of breed, Aside, *The Common Law Origins of the Infield Fly Rule*, 123 U. PA. L. REV. 1474 (1975).

⁸ See, e.g., TOBIAS MOSKOWITZ & L. JON WERTHEIM, *SCORECASTING: THE HIDDEN INFLUENCES BEHIND HOW SPORTS ARE PLAYED AND GAMES ARE WON* (2011) (exploring, among other questions, the magnitude and causes of home-field advantage, and whether teams and players really ever are “in the zone”); Joseph Price and Justin Wolfers, *Racial Discrimination Among NBA Referees*, *QUARTERLY JOURNAL OF ECONOMICS*, Oct.-Dec. 2010, at 1859-1887 (using econometrics to assess racial discrimination by NBA referees); Ignacio Palacios-Huerta, *Professionals Play Minimax*, *REVIEW OF ECONOMIC STUDIES*, Apr. 2003, at 395-415 (testing the Minimax theorem using data from penalty shoot-outs in soccer).

students of the domain of inquiry that we may denominate the jurisprudence of sport.

* * *

Although newcomers to the debates over instant replay might suppose that the issue presents a single question—whether to employ instant replay or not—it is helpful to distinguish at least three sets of questions: (1) What calls should be reviewable?; (2) what should be the procedures be for implementing review?; and (3) what should be the standard of review?

Particularistic answers to the first question must be sport-specific simply because each sport has its own rules and therefore its own calls (pass interference, strike, offside, traveling, leg before wicket, etc.). But necessarily sport-specific particularistic answers might possibly follow from a more general solution that could apply across sports—for example, from a principle that all and only the “game-changing calls” should be reviewable, or that review should be available for all except the “judgment calls.” General solutions like these are conceivable, but do not seem promising on reflection—partly because the categories of “game-changing calls” and “judgment calls” might not withstand scrutiny,⁹ and partly

⁹ It is difficult to separate types of calls into those that are and are not “game-changing” because a particular instance of a call-type that does not usually substantially influence the game’s outcome might be momentous on that occasion. Imposition of a modest five-yard penalty for delay of game or illegal procedure, for example, might push the offense out of field goal position. Moreover, whether it does might be known only to the kicker and his coach, which is why critics have objected to the NCAA rules that allow booth review only when, among other things, the call at issue would have—in the replay official’s estimation— “a direct, competitive impact on the game.”

The distinction between “judgment calls” and “non-judgment calls” might appear more promising. Indeed, although the NFL rules specifically enumerate the reviewable calls, many commentators have described—and endorsed—the classificatory principle at work in these more general terms. The argument runs like this: Even if we should try harder to get calls right when there is a true fact of the matter to discover (like whether the ball hit the ground or crossed the plane of the end zone), there just aren’t simple truths about matters of evaluative judgment (like whether some amount of contact was excessive), so subjecting such calls to replay review would amount to substituting one official’s judgment for that of another. See, e.g., Chad M. Oldfather & Matthew M. Fernholz, *Comparative Procedure on a Sunday Afternoon: Instant Replay in the NFL as a Process of Appellate Review*, 43 Ind. L. Rev. 45, 51-52 (2009).

But matters are not quite so neat. Most importantly, many unreviewable “judgment calls” involve elements that are purely factual. Take the most significant of the unreviewable calls: pass interference. How much contact is necessary to constitute forbidden interference might well be a matter of judgment and not properly reviewed. But what if replay shows that there was no contact at all, that the receiver tripped over his own feet? A mistaken defensive pass

because, even to the extent that categories like these are coherent, the costs and benefits of review are likely to vary across sports to such a degree as to render one-size-fits-all solutions suboptimal. To focus on just one consideration, any review consumes some amount of time, and sports vary in their tolerance for delay—soccer, for example, prides itself on the continuity of its play, whereas cricket matches are cheerfully interrupted for tea. So it seems likely that some sports will find it desirable to review even some calls that aren't game-changers, and other sports will reasonably prefer not to review even some calls that are thought not to require application of an official's (subjective) "judgment." Much the same reasoning would apply to the second question too: procedures for instituting review on a given occasion must be sensitive to variations among sports on matters relevant to instant replay process.

Even if answers to the first two questions vary from sport to sport, many people seem to think that the answer to the third is fairly general—not necessarily that there is one optimal review standard for all sports that institute instant replay, but that there is one right answer for many. The conventional view is that on-field calls should be entrenched against reversal such that they may be overturned only when the replay evidence is "indisputable" (in the language of the NFL, NCAA football, and the CFL) or "conclusive" (e.g., NBA and Test cricket). Put another way, it is standardly accepted that if the reviewing official has any doubt as to the correctness of the initial call, he should be instructed to let it stand—even if he considers it very probably wrong.

Precisely because the "indisputable visual evidence" (IVE) standard is widely embraced as the appropriate standard of review across disparate sports,¹⁰

interference call in such circumstances seems like just the sort of thing that review is well suited for.

Just as interestingly, reviewable calls can contain unnoticed elements of judgment. Consider a potentially game-altering play in the Giants' 38-45 loss to the Eagles in Week 14 of the 2009 season. After scrambling for a 15-yard gain to escape a sack, Giants QB Eli Manning fell to the ground and lost the football. The Eagles recovered and were awarded possession. Calls of fumble and change of possession are reviewable to determine whether the ball carrier was down by contact before losing control of the ball, and the Giants challenged. Replay showed indisputably that Manning's knee was down before he lost the ball. So ordinarily the call would be reversed. But while it was clear that Manning was down, it wasn't as clear that he was down *by contact*: Eagles' defensive tackle Brodrick Bunkley had grabbed a piece of Manning's jersey as he escaped the pocket, but Manning took a few steps before hitting the turf, making it unclear whether the defensive contact had caused Manning's contact with the ground. If it had, there was no fumble; if it hadn't, there was. In the event, referee John Parry upheld the call on the field, but neither he nor the broadcast announcers suggested that this was a judgment call, hence not reviewable.

¹⁰ See e.g., Gabrielle Marcotti, *Further Review on Instant Replay*, SPORTS ILLUSTRATED, Sept. 25, 2008, available at

that is the focus of this essay. Concentrating on football, I will argue that conventional wisdom in favor of IVE likely rests upon mistaken premises, and that settled practices should probably be changed. The analysis unfolds over four parts.

Part I presents a brief history of instant replay review in the NFL. Part II introduces the *prima facie* case in favor of a less demanding standard than IVE. It is straightforward. Because the reason to institute instant replay review in the first place is to correct officiating errors, the various features of the replay system—including the standard of review—should be designed to minimize final errors. Drawing widely from actual experience with the standard in professional football and from longstanding practices in the law, it argues that a standard, like IVE, that entrenches the initial call does not serve this interest, and that the goal of error-correction would be better served by employing a less demanding standard, such as *de novo* review.

Of course, minimizing officiating errors cannot be the only criterion by which to evaluate standards of review: even if a *de novo* standard of review would yield fewer final errors than does IVE, it might nonetheless be suboptimal if it proved too costly in other respects. Accordingly, Part III isolates and assesses possible counterarguments to the basic case against IVE presented in Part II. It finds unpersuasive the most prominent arguments against a less stringent standard of review—that such a standard would intolerably increase reviewing delays or would unduly threaten respect for the officials. However it also identifies other reasons to disfavor *de novo* review—reasons not already well developed in the literature—that are likely to resonate with some readers. It concludes, therefore, that no standard of review for the use of instant replay in football obviously dominates the alternatives given reasonably divergent empirical assumptions and evaluative judgments. Fortunately, the ambivalence of its conclusion notwithstanding, the disparate analyses of Part III allow for a much clearer-eyed assessment of the status quo, and of alternative proposals for the use of instant replay review, than was possible previously. Part IV makes good on this claim by presenting three novel proposals for revising the football replay system to accommodate the different perspectives on the value and costs of error-correction that emerge from Part III.

Because this (partial) roadmap might convey the impression that the trip ahead will be of interest (at most) only to readers who are glued to their sets every Sunday in the fall, I should assure you that it is not quite so parochial. First,

http://sportsillustrated.cnn.com/2008/writers/gabriele_marcotti/09/25/replay/ (arguing that soccer should institute instant replay and that the appropriate standard of review is “simple. Just adopt the NFL language.”).

although the express focus throughout will be on American football to keep things manageable, the argument will draw occasionally from practices in other sports, baseball most particularly. And just as the discussion will sometimes draw from other sports, so too will it bear lessons for them: implications for all sports that presently use, or contemplate introducing, instant replay will rarely lie very far below the surface. Second, we cannot accomplish the narrow goals specified without investigating some issues whose philosophical or jurisprudential relevance will be apparent even to readers uninterested in sports.¹¹ Trying to figure out the optimal standard of review for instant replay in football will require that we confront difficult questions on such disparate matters as the relevance of both desert and entitlement to justice, the possible difference between mistakes and errors, and the likely shape and strength of varieties of cognitive bias.

Furthermore and finally, while the first four parts of the essay are principally in service of the idea that law has something to teach sports, law and lawyers don't have all the wisdom. The jurisprudence of sport, after all, is predicated on the belief that sporting practices and norms can teach lessons for ordinary legal systems just as surely as the other way around. Part V illustrates this proposition by drawing from replay practices in professional and collegiate football reason to believe that the Anglo-American criminal trial system might be profitably reformed to accommodate three verdicts, not two.

I. UNDER REVIEW: VIDEO REPLAY IN THE NFL

Riding Terry Bradshaw's arm and their fearsome Steel Curtain defense, the Pittsburgh Steelers defeated the overmatched Los Angeles Rams 31-19 in Super Bowl XIV to win an unprecedented fourth championship. Yet some observers (especially on the Gulf Coast) thought that the Steelers' victory was tainted. Two weeks earlier, they had beaten the Houston Oilers in the 1980 AFC Championship Game by the seemingly lopsided score of 27-13. But the final score was misleading. With the Steelers up 17-10 in the closing seconds of the third quarter, Oiler quarterback Dan Pastorini hit Mike Renfro in the back of the end zone. Although Renfro appeared to have control of the ball with both feet in bounds before tumbling out of the end zone, officials ruled the pass incomplete and the Oilers had to settle for a field goal. The Steelers scored 10 unanswered points in the fourth quarter to clinch the victory and a trip to Pasadena. Television replays at the time, however, showed, clearly enough to most viewers at home, that Renfro had control before he fell out of bounds. It should have been a touchdown.

¹¹ You might think me overly optimistic in anticipating that anyone uninterested in sport would have read far enough to reach this footnote. However even legal theorists have mothers.

The “Mike Renfro Game” had legs. Six years later, NFL owners cited it and other miscalls when voting to introduce a system of instant replay review for the 1986 season.¹² Under the inaugural system, review was entirely in the hands of a single replay official sitting in a booth who, armed only with the broadcast feeds and two VCRs, had sole authority to review a limited range of calls, mostly relating to possession (like fumbles and receptions) and boundaries (the sidelines and goal lines). A call could be reversed only if the evidence that it was incorrect was “indisputable.”

This was hardly a streamlined system, and it predictably produced lengthy delays. By 1991, replay officials reviewed over a thousand calls at an average of over three minutes each, adding nearly 15 minutes to each game.¹³ Moreover, the NFL acknowledged that fully 13% of the reversals were mistaken.¹⁴ In 1992, the owners voted to eliminate video replay.

If the abandonment of that initial system was unsurprising, its eventual return was even less so. There were murmurings, for example, when the Buffalo Bills lost a 1996 playoff game to the Jaguars, after the Jaguars recovered a fourth-quarter Jim Kelly fumble and drove for the winning field goal, while television replays revealed that Kelly had been down by contact before losing the ball. But discontent reached a boiling point two years later.

In week 13 of the 1998 season, the Bills were again victimized by a bad call in a 25-21 loss at New England, as the Patriots’ game-winning, time-expired touchdown came two plays after Patriot receiver Shawn Jefferson, clearly out of bounds, was credited with a reception for a first down on a desperation fourth-down pass. And the next week, the Seahawks lost to the New York Jets, when, with seconds remaining, quarterback Vinny Testaverde snuck in on fourth and goal from the Seattle 1 to give the Jets a thrilling 32-31 victory. Problem was, the Seattle defense had clearly stopped Testaverde a good yard from the goal line. That “phantom touchdown” cost Seattle a playoff berth and Coach Dennis Erickson his job. Then, in the first round of those playoffs, the 49ers knocked off the visiting Packers 30-27, thanks in part to the referees’ failure to call a Jerry Rice fumble on the winning drive.

Responding to season-affecting mistakes such as these, the owners voted overwhelmingly to reintroduce a new instant replay system on a trial basis to start in 1999. Rich McKay, general manager of the Tampa Bay Buccaneers and co-

¹² *What If the NFL Had Not Employed Instant Replay*, ESPN, Feb. 13, 2007, available at <http://sports.espn.go.com/nfl/news/story?id=2764251>.

¹³ Lonnie White, *Instant Replay—Football: NFL Gives It a Second Look, and Tagliabue Says New Computerized System Is Huge Improvement Over Version Used in the Late '80s*, L.A.TIMES, Sept. 12, 1999.

¹⁴ *Id.* (noting that 12 of 90 reversed calls were mistaken).

chairman of the NFL competition committee, explained the league's reasoning succinctly: "we want to get the play right."¹⁵ As the NFL's Senior Vice President for Broadcasting, Val Pinchbeck, elaborated, the league doesn't "want anyone not going to the Super Bowl because of an error that could be corrected."¹⁶

That trial system, occasionally modified, was made permanent in 2007. It differs from the system that the league had employed from 1986 through 1991 in several respects. First, it provides that many more calls would be reviewable—a list that the league has periodically expanded.¹⁷ Second, except in the final two minutes of each half, replay is initiated, not by a replay official, but by the head coaches who are permitted two challenges per game, and a third if each of the first two is successful. An unsuccessful challenge costs the challenging team a timeout. Third, the review itself is conducted, not by a booth official, but by the referee who is given access to feeds from all stadium cameras, but is permitted to view the video for only one minute.

One critical feature of the system has not changed over these many years: the standard of review. The rules on this point are very clear: "A decision will be reversed only when the Referee has indisputable visual evidence available to him that warrants the change."¹⁸

II. THE NUTSHELL CASE AGAINST IVE

The goal of instant replay is to correct officiating errors—or at least those errors that are both correctible at reasonable cost, and consequential. Because form should follow function, the instant replay system should be designed to minimize (consequential) officiating errors. More precisely, the various features of the system—including the standard of review—should be chosen to maximize error correction up to the point at which the marginal cost exceeds the marginal

¹⁵ Thomas George, *N.F.L. Backs Limited Replay After Complaints of Bad Calls*, N.Y. TIMES, Mar. 18, 1999, at A1.

¹⁶ Charean Williams, *Play it Again: The NFL's Instant Replay Isn't Always the Last Word*, ORLANDO SENTINEL (1999).

¹⁷ Reviewable calls include: all those related to whether a player or ball did or did not cross a goal line or side line; whether a pass was completed or intercepted; whether a pass was touched by an ineligible receiver or by a defender, and whether it was thrown beyond the line of scrimmage; whether a ball carrier was down by contact before losing control of the ball; whether a kick attempt was good, when not higher than the top of the uprights as it crossed the goal post; whether there were too many players on the field; and the spotting of the ball when a correct spot might determine whether the offense is awarded a first down. NFL Rules, Rule 15, § 9. The list of unreviewable calls is much longer. Some of the more common and consequential include offside, holding, pass interference, unsportsmanlike conduct, status of the clock, forward progress not relating to touchdowns or first downs, and recovery of a loose ball in the field of play.

¹⁸ NFL Rule 15 Section 9.

benefits. Call this the Pinchbeck Principle. Stripped to its essentials, the argument against IVE is that it does not satisfy this desideratum, and that a more relaxed standard of review—something like the law’s *de novo* standard of review—would.

A. *The Heart of the Argument*

“Indisputable visual evidence” is an extraordinarily stringent standard—on its face, more demanding than the “beyond a reasonable doubt” standard of proof that governs criminal trials. The intended consequence is to drastically minimize the number of times a correct initial call is reversed. A second consequence—not intended, but surely foreseeable—is to ensure that a large number of mistaken initial calls are permitted to stand uncorrected. Put another way, if the standard is complied with, correct on-field calls will be overturned very rarely. That’s good. But incorrect on-field calls will be allowed to stand very frequently. That’s bad. From an error-minimization standpoint, the only question is which effect is greater relative to a more neutral or less deferential standard of review: the reduction in erroneous reversals or the increase in erroneous affirmances?

Surely, however, this is not a difficult question. On the assumption that reviewing officials’ beliefs track reality tolerably well—or, more precisely, that they are not systematically biased in one direction or the other¹⁹—the league would minimize the combination of erroneous final calls by instructing each official to announce the ruling that he believes is more likely than not correct. Error minimization, after all, is precisely the rationale behind the law’s preponderance of the evidence—or “more likely than not”—standard of proof.²⁰

¹⁹ This is a critical assumption that, although generally reasonable, isn’t always. Consider three-card monte. What makes this con successful is that the mark’s beliefs are poor guides to truth. If you’re looking for, say, the red ace, you might do better *not* picking the card in which you have most confidence. The possibility of similar and systematic mismatch between facts and beliefs is something a legal system should take into account. For example, controlled experiments repeatedly demonstrate that people place vastly too much credence in eyewitness testimony—a proposition that the large number of cases of DNA exoneration vividly bolster. So when only one party to a lawsuit introduces such testimony, factfinders are likely to be more confident in that party’s case than is warranted. Or suppose that jurors systematically overweight the testimony of police officers. If so, then use of the preponderance standard in criminal trials might produce more false positives than would an epistemically unbiased factfinder. A standard of proof greater than preponderance could help counter this bias, and actually increase total accuracy. Unsystematic biases—biases that sometimes favor one party, sometimes the other—might wash out in the long run. But when a legal system is infected by recurring and nonrandom biases, then the more-likely-than-not standard of proof might not minimize total errors relative to a heightened (or lowered) standard designed to offset the bias.

²⁰ See, e.g., DH Kaye, *The Error of Equal Error Rates*, 1 Law, Prob. & Risk 3 (2002) (discussing and criticizing the competing hypothesis that preponderance of the evidence equally allocates errors).

Much as unbiased jurors will, in the aggregate, minimize the sum of false negatives and false positives by announcing verdicts in accordance with the facts they believe more likely than not correct, so too would replay officials minimize the sum of erroneous affirmances and erroneous reversals by announcing the rulings in which they have the greater confidence. If, as would appear, replay officials have good access to the facts, strongly deferential review all but ensures more errors than necessary.

B. The Saints Go Marching In (An Illustration)

The 2010 NFC Championship Game came down to this: a 40-yard field goal attempt by Saints' kicker, Garrett Hartley, less than five minutes into overtime. A successful try would earn New Orleans a trip to Miami for the first-ever Super Bowl in the 'Aints lackluster 43-year history; a miss would give the visiting Minnesota Vikings possession at their own 30-yard line and another chance to break the hearts of the Saints' long-suffering fans. Minnesota called time to ice the third-year kicker, but Hartley was unnerved, splitting the uprights for a thrilling 31-28 Saints' victory and a date against the favored AFC Champion Indianapolis Colts, whom they would defeat two weeks later.

Although the contest had seen more than its fair share of big plays and controversial calls, the game-winning drive was unusually event-filled, with four close and questionable rulings occurring over an extraordinary span of just five plays.²¹ Three of the calls, all in the Saints' favor, were particularly close: a first-down call on a fourth-and-one leap over the pile by Saints' running back Pierre Thomas; a dubious pass interference call against Minnesota linebacker Ben Leber that gave the Saints a twelve-yard gain; and a reception by wideout Robert Meachem—for another twelve-yard pick up—on a low Drew Brees delivery over the middle.

By rule, the pass interference call was unreviewable, but replay officials did review the Thomas first down and the Meachem reception, eventually upholding both close calls for want of the “indisputable visual evidence” required for reversal. Meachem's reception was an especially bitter pill for Vikings fans, for the ball clearly did hit the ground. Possibly, Mechem maintained possession, but most observers didn't think so. The consensus was that the call on the field

²¹ The game's conclusion was noteworthy for another reason too. It prompted the NFL to change its overtime rules to reduce the influence of luck in calling the coin toss by providing that a team cannot win a playoff game in overtime just by scoring a field goal on the first overtime possession. See generally David Fleming, *The Making of NFL Rule 16*, ESPN The Magazine, Dec. 27, 2010, at 74. The new rule provides that “If the team that possesses the ball first scores a field goal on its initial possession, the other team (Team A) shall have the opportunity to possess the ball. If Team A scores a touchdown on its possession it is the winner. If the score is tied after Team A's possession, the team next scoring by any method shall be the winner.” NFL Rule 16.

was mistaken: should not have been ruled a catch. For example, under the headline, “Did Officials Bungle the End Game for Vikings?” the *New York Times* sports blog observed that “Meachem bobbled a catch as he fell to the ground at the Vikings’ 22. The ball appeared to hit the ground and move.”²²

But if, as dominant opinion would have it, the call on the field was mistaken, most commentators believe that referee Pete Morelli’s refusal to overturn the call was correct too. Analysts at NBC Sports “agreed with the decision. Though it appears that the call was wrong, the ‘100 drunks in a bar’ standard requires clear evidence to overturn the real-time ruling. And clear evidence was not available.”²³ NFL VP of officiating, Mike Pereira, made precisely that point when acknowledging that the call could have gone either way while insisting that the refusal to overturn the call was, therefore, necessarily correct. Under league rules, he explained, referees are not to overturn on-field calls “unless [they] had 150% concrete visual evidence that there was no doubt in anyone’s mind” that the call was mistaken.²⁴

Notice, then, that replay permitted this call to be reviewed; the call was probably mistaken; and IVE prevented it from being corrected. So, we might reasonably ask: Was it a consequential mistake? What would have happened had this apparent mistake been corrected?

Of course, that’s impossible to answer for sure. But we can reason through some probabilities. If the referee had ruled, after review, that Meachem had not retained possession, New Orleans would have faced third and fifteen at the Minnesota 34. The Saints had converted only 3 of 12 third downs the entire game, and their chances at that distance would have been dim.²⁵ So the likelihood is that (had this probably-mistaken call been corrected) Saints coach Sean Payton would soon have faced a difficult choice: go for it on fourth down, attempt a field goal, or punt. No doubt his choice among these options would have been affected

²² *Did Officials Bungle the End Game for Vikings?* N.Y. TIMES, Jan. 24, 2010, available at <http://fifthdown.blogs.nytimes.com/2010/01/24/did-officials-bungle-the-end-game-for-vikings/>.

²³ Gregg Rosenthal, *Pereira Talks About Controversial Calls in Saints-Vikings Overtime*, ProFootballTalk.com, Jan. 27, 2010, available at <http://profootballtalk.nbcsports.com/2010/01/27/pereira-talks-about-controversial-calls-in-saints-vikings-overtime/>.

²⁴ *Championship Playoffs Official Review Bonus Coverage (video)*, NFL.com, Jan. 27, 2010 available at <http://www.nfl.com/videos/nfl-playoffs/09000d5d816037d3/Championship-Playoffs-official-review-bonus-coverage>

²⁵ Over the course of the 2009 season, offenses facing third down and at least 8 yards to go had converted just under 24% of the time. *Inside 3rd Down Run v. Pass Success*, Stampede Blue, Jun. 24, 2010, available at <http://www.stampedeblue.com/2010/6/24/1533609/inside-3rd-down-run-vs-pass>. I haven’t calculated the conversion percentage for third and fifteen, but we can bet that it was noticeably lower.

by how many yards, if any, New Orleans had gained on third down. But neither of the first two options promised great hope for success: an incompleteness on third, for example, would have left Hartley with a 51-yard field goal attempt, well beyond his career long.

So the chances are good that, one way or another, the ball would have been delivered once more into the hands of the powerful Minnesota offense led by Brett Favre—the 2009 miracle man who, in his first season in purple and gold was coming off perhaps the best season of his 19-year Hall of Fame career, not the 2010 version who appeared to be auditioning for the *Walking Dead*. And had Favre again seen the ball, Vikings fans would have had reason for optimism. Their high-octane offense had already shredded the Saints' D for 475 yards, and their kicker, the 13-year veteran Ryan Longwell, had been money all season long: his 92.9% field-goal completion percentage—including two for two at fifty yards and beyond—was second best in the league.

In short, the reception call was probably mistaken, it was correctly upheld under the governing IVE standard but probably should and would have been overturned under *de novo* review, and the mistake might well have made the difference for the Vikings between going home and going to Miami. To repeat, we can't know for sure what would have happened had this probably mistaken call been reversed—whether the Vikings would have gained possession again and, if so, what they would have done with it. But the thought that it might have made all the difference is no mere fantasy. Recall the rationale for employing instant replay in the first place: to ensure, to the best of the league's ability, that no team doesn't go "to the Super Bowl because of an error that could be corrected." If so, then the 2010 NFC Championship game suggests that the standard of review is ill-chosen: it should be *de novo*, not IVE.

C. Summary

Epistemic standards that provide that an initial call cannot be corrected after instant replay review unless the reviewing official deems the evidence of error "indisputable" or "conclusive" strongly entrench initial calls against reversal. The argument against strong entrenchment—and, equivalently, the argument in favor of something like *de novo* review—can be parsed as follows:

1. The NFL should employ that standard of review that would maximize the correction of consequential errors except insofar as correcting the marginal error is outweighed by marginal costs of the system. (The Pinchbeck Principle)
2. *De novo* review corrects more (consequential) errors than does IVE. (The error-minimization premise)

3. *De novo* review does not incur more costs than IVE—or not enough to outweigh the increased benefits in the coin of error correction. (The not-prohibitively-costly premise)

4. Therefore, *de novo* review is preferable to IVE.

III. IN DEFENSE OF IVE

I am aware of no systematic defense of the IVE standard. To my knowledge, the NFL itself has never publicly explained the reasoning behind its choice of standard, and because the puzzle has received so little attention from fans and commentators, it is hard to be sure of all that might be said in its favor. That said, I have seen only a handful of arguments in favor of the IVE standard. And most, unfortunately, are blog posts.²⁶ Some of these arguments are implausible on their face;²⁷ others are frustratingly enigmatic.²⁸ My goal in this

²⁶ Toward the end of 2009, Duke law professor Joseph Blocher (responding to an article that had discussed the IVE standard approvingly, Oldfather & Fernholz, *supra* note 9) questioned the NFL's IVE standard of review in a post to Prawfsblawg. Joseph Blocher, *Why Aren't Instant Replays Reviewed De Novo*, PrawfsBlog, Dec. 1, 2009, available at <http://prawfsblawg.blogs.com/prawfsblawg/2009/12/why-arent-instant-replays-reviewed-de-novo.html>. That post was picked up by several other bloggers, including George Mason's Ilya Somin. Ilya Somin, *Why Instant Replay Should Be Like De Novo Appellate Review*, The Volokh Conspiracy, Dec. 2, 2009, available at <http://volokh.com/2009/12/02/why-instant-replay-should-be-like-de-novo-appellate-review/>. Having already completed a draft of this chapter, I wrote a short piece for Slate extending Blocher's and Somin's arguments. Mitchell Berman, *After Further Review*, Slate, Dec. 22, 2009, available at <http://www.slate.com/id/2239018/>. These postings, and others, provoked a large number of rejoinders. I will be quoting from some of those comments.

Blocher was not the first to recommend that the NFL adopt a less demanding standard of review. Nearly a decade earlier, a short piece appeared in the Vermont Law Review proposing that “manifest weight of the evidence” would be a preferable standard. Jack Achiezer Guggenheim, *Blowing the Whistle on the NFL's New Instant Replay Rule: Indisputable Visual Evidence and a Recommended “Appellate” Model*, 24 Vt. L. Rev. 567 (2000). That analysis, however, was almost entirely based on a peculiar premise. Observing that, midway through the 1999 season, nearly 30 percent of challenged calls had been reversed, Guggenheim concluded that the reviewing referees must be applying something less stringent than IVE in practice. Yet, he further noted, league executives expressed satisfaction with the system, so they must want a 30% reversal rate. The “manifest weight of the evidence” standard, he concluded, would capture existing, and seemingly desired, practice more faithfully than does the IVE. *Id.* at 569-71. The fly in the ointment, though, is Guggenheim's assumption that the bare reversal rate can tell us anything at all about the standard of review being applied. That the referees were not applying the IVE standard would have been a sensible inference if the calls they had been reversing at a 30% clip had been selected at random, but not if, as was the case, the reviewed calls were selected from a vastly larger pool by coaches and replay officials who knew what the applicable standard would be.

²⁷ See, e.g., Posting of bsharpoflat to Slate, <http://fray.slate.com/discuss/forums/thread/3534251.aspx?ArticleID=2239018> (contending that “the essence of sports spectatorship demands that the referee's onfield decision take precedence

Part is to disambiguate the arguments for IVE and to carefully assess them. This will require some sympathetic reconstruction. It will also demand more argumentative care than “serious” people are usually inclined to devote to the subject.²⁹ Indeed, some readers will likely find my efforts here perverse on the grounds that football is, after all, “only a game.”

But, of course, my campaign for a rigorous jurisprudence of sport is predicated on a contrary assumption—namely, that sport warrants and repays exacting analysis. This is partly because sports present intellectual problems that are intrinsically interesting and challenging, and partly because those problems are often structurally analogous, sometimes structurally identical, to problems that arise in domains of life of greater intrinsic importance. So solutions in this humble domain will often bear fruit for loftier realms as well. In this way, and much like the space program, the jurisprudence of sport promises positive spillover effects.

All sound arguments for IVE should be formulable as challenges to one or another step in the argument for *de novo* review. Because the conclusion seems clearly to follow from the premises, the objections can be grouped as challenges to one or another of the first three premises. Section III.A considers a challenge to premise 2—i.e., that the IVE results in more total mistaken calls than would a lower standard of review. On this competing view, there is no need to engage in the careful and contestable balancing of pros and cons that premise 3 aims to capture because the key consideration that I have claimed to weigh against the NFL’s chosen standard is simply fallacious. This argument, I conclude, is meritless. Two additional arguments acknowledge that the IVE produces more total errors than would a lower standard but identify countervailing considerations that, their proponents believe, justify the more demanding standard all things considered: that the IVE is reasonably necessary to reduce the delay that use of replay technology introduces; and that the IVE protects against corrupt officials.

over a video-shows-most-likely-correct decision”). Non-deferential review might be a bad idea, but the notion that it would flout essential demands of sports spectatorship is hard to take seriously.

²⁸ See, e.g., Posting of Allen to Prawfsblawg, <http://prawfsblawg.blogs.com/prawfsblawg/2009/12/why-arent-instant-replays-reviewed-de-novo.html#comments>. (“To me the current system makes sense. Deferential review because there are instances where the officials see things the cameras can’t. Replays overrule the officials where the cameras see the same things, and the result is different.”)

²⁹ As the venerated English sportswriter Simon Barnes observed, “One thing that sex and sport have in common is that stupid people like them both. One important difference is that clever people can enjoy sex—and can say that they enjoy sex—without forfeiting their right to be considered clever. However, a clever person who claims to enjoy sport will be considered less clever as a result.” SIMON BARNES, *THE MEANING OF SPORT* 48 (2006).

These arguments, in other words, deny premise 3. They are scrutinized and found wanting in Section III.B.

The most interesting and complex arguments for IVE take on premise 1, the Pinchbeck Principle. The first argument (Section III.C.1) maintains (on two distinct grounds) that the league's goal should not be to minimize officiating errors simpliciter, but to minimize *weighted* officiating errors, where errors produced by mistakenly reversing an initially correct call are worse than errors produced by affirming an initially mistaken call. The second argument (Section III.C.2) would reinterpret the Pinchbeck Principle. On this view, the league's goal should not be to correct officiating *mistakes*. Instead, it should be to correct officiating *blunders*. And while *de novo* review might correct more mistakes, the argument continues, it does not correct more blunders.

A. Challenging Premise 2: The IVE Minimizes Total Officiating Mistakes

I have said that the IVE yields more total errors than would a less deferential standard of review on two assumptions: that replay officials administering a heightened standard do in fact let stand some calls they believe are more likely than not mistaken; and that when a replay official believes, based on all the evidence available to him, that an on-field call was more likely than not mistaken, he is more likely than not correct. Critics have taken issue with both premises.

I don't know how best to respond to the first objection—that, as one blog comment put it, “IVE seems to be regularly disregarded and *de novo* review regularly applied, so any call to go to a *de novo* standard would essentially only codify what the refs are already doing.”³⁰ This is conceivable, though I'm fairly confident that it's a minority view (Vikings fans don't share it), and surely not one that the NFL itself would advance. More interesting is the second objection—that the refs do apply a highly deferential standard (true IVE, or something reasonably close to it), and that doing so is in fact the best way to minimize total mistaken calls.

People who make this argument often start by emphasizing that the IVE standard is of review, not of proof. In the law, factual decisions by juries and trial judges are generally reviewed on appeal under the deferential “clearly erroneous” standard. The rationale, of course, is not that the trial judge “knows more than his loftier brothers; rather, he sees more and senses more.”³¹ He has access to informal types of evidence—things like witness demeanor and the curious gaps

³⁰ Posting of dsimon to Slate available at <http://fray.slate.com/discuss/forums/thread/3533936.aspx?ArticleID=2239018>

³¹ Maurice Rosenberg, *Judicial Discretion of the Trial Court, Viewed from Above*, 22 *Syr. L. Rev.* 635, 663 (1971).

and pauses in an attorney's questioning of a witness—that appear pallidly, if at all, in the formal written record that appellate courts review. Insofar as an appellate tribunal might fail to appreciate the extent of the initial decisionmaker's comparative epistemic advantage, a *de novo* standard of appellate review is likely to increase rather than decrease factfinding errors. Just as the law's "clearly erroneous" standard is designed to insulate an initial finding from *erroneous* reversal, so too, the argument goes, does football's IVE standard protect initial calls precisely because the on-field official is more likely to be right than is the reviewing official: "the referee (or jury) is most likely to get the call right because he has seen the actual evidence first-hand, while the reviewing official (or appellate court) is limited to seeing what was captured on camera (or the court transcript)."³²

I don't think so, for an appellate judge stands in a very different relationship to facts that are disputed at trial than does a replay official to facts that transpire on the gridiron. The trial judge and the jury *do* see more and sense more when it comes to evaluating the testimony of witnesses. But it's hard to swallow that, in general, a single field official sees more through the single pair of eyes with which nature has endowed him than does the replay official who views the field of play through, as it were, the lenses of many cameras. And not only does the replay official have access to multiple camera angles, he views the images from those cameras in high definition at super slo-motion. Given the epistemic aids that the replay official enjoys and that the on-field official lacks, it is implausible that the latter's greater physical proximity to the event would make a decisive difference. Indeed, most sports officials know that video review allows for more accurate factual determinations than can be made in real time, which is precisely why some resist instant replay. As Italy's Pierluigi Collina, the top-ranked soccer referee in history, complained: "Referees are blamed because they are unable to see what can be shown by technology that has made giant steps forward. No matter how prepared referees are and how hard they work, they will never be able to compete with technology."³³ So the suggestion that *in general* the single on-field official has superior access to the facts than does the replay official strikes me as patently silly. If all we are concerned about is improving

³² Posting of matchcliff to Slate available at <http://fray.slate.com/discuss/forums/thread/3539860.aspx?ArticleID=2239018>

³³ Ref Collina Clarifies Comments on Video Officials, Rediff.com, Dec. 30, 2009, available at <http://sports.rediff.com/report/2009/dec/30/referee-collina-clarifies-comments-on-video-officials.htm>. For the ranking of referees by the International Federation of Football History and Statistics, see "The World's Best Referee of the Quarter of the Century," International Federation of Football History and Statistics, available at <http://www.iffhs.de/?20e43c03f32b00f31c13f32b17f7370eff3702bb1c2bbb6e08>.

accuracy—getting the *final* call right—then a *de novo* standard of review is more sensible.³⁴

None of this is to deny, however, that, even with the benefit of all this technology, there are some cases in which the on-field official is better situated to make the correct call—namely, when he had a particularly fortuitous visual angle that none of the cameras captured, as might be the case on goal line plays if there is no camera on the line. So let me clarify one thing. The *de novo* standard of review is sometimes described as a directive to the reviewing court to ignore or disregard the decision of the lower court—to proceed as though writing on a clean slate.³⁵ Consistent with this description, you might think that were a referee to review the play afresh, he must ignore the initial call entirely, trying to expunge from his mind any knowledge of it. That is not my suggestion. The on-field official is analogous to a judge in some respects. But he's analogous to other legal actors too; referees and umpires are sometimes described as a combination of judge and police.³⁶ For present purposes, though, it's more helpful to think of him as a third familiar player in a legal drama: the eyewitness.

Were a replay official to review plays *de novo*, then, far from striving to wipe from his mind the initial on-field call, he should respect it as the testimony of an eyewitness. Like all eyewitness testimony, sometimes it's really good evidence, sometimes not. But anytime the video evidence leaves a replay official close to equipoise, he'd be wise, even when applying *de novo* review, to go with the on-field call. This wouldn't be a matter of deferring to a judgment that departs from his own, but of reaching a judgment based on all the available

³⁴ One commentator has opined that “[a] *de novo* standard and the absence of deference it entails would suggest that there is little confidence in on-site officials’ ability to generally make the right call.” Guggenheim, *supra* note ___, at 578. Not so. A *de novo* standard suggests (indeed, assumes) only that the on-site official does not generally have *greater* ability to make the right call than does the replay official—although, as the text goes on to explain, such a standard could and should be used in a way that recognizes that the on-site official will *sometimes* have greater ability to make the right call than would the reviewing official. Ironically, if any standard suggests a lack of confidence in officials’ ability, it’s the IVE, which suggests little confidence in the ability of those reviewing the play to make the right call notwithstanding the patent epistemic advantages they enjoy.

³⁵ See, e.g., *Mayer v. Montgomery County*, 794 A.2d 704, 716 (Md. App. 2002) (“A *de novo* proceeding is one that starts fresh, on a clean slate, without regard to prior proceedings and determinations. A true trial *de novo* ... puts all parties back at ‘square one’ to begin again *just as if the adjudication appealed from had never occurred.*”) (internal citation and quotation omitted; emphasis added).

³⁶ See, e.g., Paul Finkelman, *Baseball and the Rule of Law Revisited*, 25 Thom. Jeff. L. Rev. 17, 25 (2002).

evidence—including the contemporaneous report of a well-positioned witness to the event.³⁷

B. Challenging Premise 3: A De Novo Standard Imposes More Costs than IVE

The third premise in the argument against entrenchment is, thus far, the least well supported. Indeed, the claim that a *de novo* standard would not meaningfully increase the costs of review has to this point just been asserted, not argued for. My strategy, following a natural enough dialectic, is to consider contrary arguments here, as challenges to the *prima facie* case against IVE. Some considerations in favor of IVE—those that focus on reversal as itself a cost—could be couched as arguments either that *de novo* review imposes greater overall costs than IVE or that *de novo* review does not minimize total weighted errors. I will defer discussion of those arguments to the next section. Here we assess purported reasons *not* linked to the supposedly greater costs flowing from each act of reversal why the costs of a reply system increase as the standard of review becomes less stringent.

1. *De novo review produces too much delay.* The introduction of video replay slows down the game. No doubt about it. That's one reason to eliminate instant replay entirely. Now, the NFL power structure doesn't think it's good enough on balance, and I happen to agree. But if you think otherwise, and assuming your view isn't based on a mistake about the empirics—say, about how much time video replay adds—then I have nothing to say to persuade you

³⁷ Compare Judge Jack Weinstein's observation that “‘*de novo*,’ while reflecting the usual Latin obscurantism of the law, does not really mean the judge will decide as on a clean slate without considering the decision of the administrator (or on appeal the decision also of the trial court). Even the most egocentric federal judge applying *de novo* review will give *some weight* to what the presumably more expert or experienced plan administrator actually did, even if that weight is applied *sub rosa* or subconsciously.” *Mood v. Prudential Ins. Co.*, 379 F. Supp.2d 267, 271 (EDNY 2005). Consider too debates over the proper review of patent claims construction after the Supreme Court's decision in *Markman v. Westview Instruments Inc.*, 517 U.S. 370 (1996). See, e.g., *Cybor Corp. v. FAS Technologies, Inc.*, 138 F.3d 1448, 1462 (Fed. Cir. 1998) (en banc) (Plager, J., concurring) (“Though we review [the] record ‘*de novo*,’ meaning without applying a formally deferential standard of review, common sense dictates that the trial judge's view will carry weight.”); *id.* at 1463 (Bryson, J., concurring) (“[O]ur adoption of the rule that claim construction is an issue of law does not mean that we intend to disregard the work done by district courts in claim construction or that we will give no weight to a district court's conclusion as to claim construction, no matter how the court may have reached that conclusion.... [W]e approach the legal issue of claim construction recognizing that with respect to certain aspects of the task, the district court may be better situated than we are, and that as to those aspects we should be cautious about substituting our judgment for that of the district court.”).

I am open to the possibility that, for this reason, a referee engaged in *de novo* review might properly have upheld the on-field ruling that Meachem had made a clean reception in overtime of the 2010 NFC Championship Game.

otherwise. Arguments that could decisively adjudicate this particular tension between the competing values of accuracy and finality are not in the cards.

But the issue isn't whether to implement video replay or not, it's what the standard of review should be once it is implemented. Total delay is a function of two variables: the number of challenges and the time taken to review each one. So a heightened standard like IVE might conceivably speed up the game—or, put better, reduce the delays that video replay introduces—by addressing either of these factors. It might deter challenges, or it might reduce the time for review once a challenge is initiated (or it might do both).³⁸

We don't want challenges because they disrupt the flow of the game, making it less aesthetically pleasing and trying fans' patience. True, we might suppose that, once a call has been challenged, we want to get it right. But we employ a heightened standard as a way to reduce the challenges in the first place, for a coach is less likely to challenge a call, thereby disrupting play, to the extent he thinks his challenge will fail (at the cost of a timeout). In a sense, then, the IVE could be understood as a type of threat justified principally on grounds of deterrence: even if *carrying out* the threat (that challenged calls will be reviewed under a highly deferential standard instead of under a standard calculated to maximize accuracy) is costly, the policy might nonetheless be justified all things considered by the good that the threat produces—invisibly, as it were—by preventing disruptions that would otherwise occur.

Although this chain of reasoning does not provide a good *explanation* for the NFL's initial adoption of IVE,³⁹ no doubt the standard of review should affect the number of challenges in just this way were coaches not limited in the number of challenges they could make. But they *are* limited—to a paltry two challenges per game (and a third if the first two succeed). We need to know, then, whether, given the cap, coaches would challenge more plays were the standard *de novo* and, if so, what the aggregate increase in review-associated delay would be.

For at least two reasons, challenges and delays might be expected to increase under a less strict standard of review notwithstanding the existing limit on challenges. First, the fact is that coaches do not use up all their challenges under the present system. Since 2004, the number of coaches' challenges made

³⁸ Although delay is one of the most common arguments given by defenders of a heightened standard of review, *see, e.g.*, Oldfather & Fernholz, *supra* note 9, at 63, these two mechanisms are rarely distinguished.

³⁹ The IVE standard was in place when the NFL first implemented replay review back in 1986 even though review was initiated by the replay official and not by the head coaches. While I do not believe that a justification for a practice is discredited merely because it arises post hoc, some might think that the fact that this rationale wasn't a plausible explanation licenses skepticism about its force.

each season has remained fairly constant at nearly 1 per game (not per team).⁴⁰ Under a significantly more lenient standard, however, we might reasonably expect coaches to come close to exhausting their allowed challenges, yielding perhaps as many as three more reviews per game,⁴¹ and conceivably as many as five additional reviews if two successful challenges still earned coaches a third. Second, perhaps we should not treat the existence of a hard ceiling as a given. No doubt the league could limit coaches to two challenges per game (and three if the first two were successful) no matter what the standard of review. But pressure to increase the number of allowed challenges might increase as the reviewing standard becomes less demanding. Fans often bristle when, because it has exhausted its challenges or its time outs, their team is unable to challenge a call that appears ripe for reversal, and there will be many more such instances the lower the standard of review. So a relaxation of the standard of review from IVE to *de novo* might bring within its train an increase in the number of challenges each team is permitted.

I am not impressed by the second argument. Any pressure to give coaches more challenges is resistible. If, in the event, it isn't resisted, that will bespeak a judgment that correcting additional errors is worth the price in somewhat greater delay.

The first argument, however, cannot be so easily dismissed: almost certainly, more calls *would* be challenged were the league to lower the standard of review while preserving the basics of its current replay system. Still, the league would have several means available to mitigate the impact on total game time. First, if least significantly, the league could repeal the rule that awards a team a third challenge when its first two are successful.⁴² Second, delay would decline if booth challenges were abandoned, and a lower standard of review would make that salutary change more palatable. Since 2004, two replay trends stand out: first, the number of booth-initiated reviews has steadily increased while the number of coaches' challenges has remained nearly constant at 300 per season; second, while the coaches have been increasingly successful with their challenges, the success rate for booth-initiated review has fluctuated from a low of 22% in

⁴⁰ See statistics at nflgisis.com.

⁴¹ Then again, because "timeouts are precious," see Dave Anderson, *Replay Instant Replay*, NY TIMES, Marc. 11, 1997, available at <http://www.nytimes.com/1997/03/11/sports/replay-instant-replay.html?pagewanted=2&src=pm> (quoting Bill Parcells), coaches would still use their challenges cautiously even if the likelihood of success were somewhat greater than it is now.

⁴² As I will explain later, I think that rule is appropriate—indeed, probably too grudging—when the standard is IVE. See *infra* Section IV.C. But a hard cap is much less objectionable when the standard is *de novo*. This is essentially for reasons that track the mistake/blunder distinction to be introduced *infra* Section III.C.2.

2004 to a high of 35.64% in 2007. In 2010, nearly 44% of all coaches' challenges were successful compared to a mere 25% for booth-initiated reviews.⁴³ What all this means, of course, is that booth-initiated reviews contribute disproportionately to total replay-associated delay while offering considerably less in the way of error-correction.

At the same time, booth reviews often interfere significantly with the game's competitive dynamic because they stop the clock during a period of the game in which the teams are particularly likely to have asymmetric interests in stoppage—either the offense benefits because it's rushing against the clock, or the defense welcomes stoppage so it can catch its breath and make desired substitutions against a hurry-up offense. Against all this, booth review is justified on the grounds that, during the high-pressured end of halves, coaches shouldn't face the added burden of having to decide quickly, before their assistants have reviewed the play from the broadcast feeds, whether to challenge close calls. That's a defensible judgment. However, coaches will need less time to decide whether to pull the trigger if they face a less deferential standard of review.

Other reasons to worry less about added delay derive from a consideration already mentioned: total delay is a function not only of the number of calls reviewed, but also the average duration of each review. Presumably, the replay official will frequently exhaust the limited time allotted him no matter what the standard of review is. So a choice between these two standards will affect review time in only two circumstances. If the reviewing official comes quickly to conclude with confidence that, regardless of whether the initial call was correct, no more careful attention to additional camera angles will be able to establish that that call was indisputably mistaken, then the IVE standard would permit him to wrap things up faster than would a *de novo* standard. On the other hand, if he promptly determines that the initial call was wrong more likely than not, and would use additional time only to mine for the shot that would seal the deal—that is, to assure himself that the incorrectness of the on-field call is indisputable—then the *de novo* standard would speed things up. My sense, as a reasonably experienced spectator, is that the latter circumstance is likely to predominate over the former. If so, we might expect a slight reduction in average time of review under *de novo* review even if the league keeps its present 60-second limit on the time a referee may watch the video. Even more importantly, the league might feel emboldened to further lower the time limit. Lastly, there are other ways too that the league could reduce the average time for review. I'll propose one in Part IV.

2. *De novo review facilitates corruption.* A second reason to believe that *de novo* review would be costlier than IVE invokes what one sports fan has

⁴³ See nflgisis.com. (Media password needed for access to database)

dubbed “the Tim Donaghy angle” after the disgraced NBA referee.⁴⁴ The theory is that “requiring a high standard of review for replay discourages refs from intentionally making wrong calls at key moments,” which in turn deters refs from agreeing to fix games.⁴⁵

That’s possible. But the argument overlooks a flip side: a high standard also reduces the likelihood that mistaken on-field calls will be reversed, thereby increasing ability of all the officials *other than the referee* to fix games. Moreover, that most subjective of officiating calls—pass interference—is also among the calls that can most influence a game’s outcome. Yet it has always been, and is likely to remain, unreviewable. (Recall the questionable defensive pass interference call against Viking linebacker Ben Leber on the Saints’ game-winning overtime drive in the 2010 NFC Championship Game. Similarly, Seattle Seahawks fans are still seething over the offensive pass interference call against wide receiver Darrell Jackson that nullified a touchdown reception in the first quarter of Super Bowl XL.) As a consequence, a corrupt field judge, side judge or back judge has about all the latitude he’d need to affect results regardless of the standard of replay review. On balance, then, the IVE would seem to trivially affect the NFL’s vulnerability to game-fixing by rogue officials.

C. Challenging Premise 1: The NFL’s Goal Is Not Mistake-Correction

The third possible defense of IVE runs through premise 1, the Pinchbeck Principle. At first blush, the Pinchbeck Principle would seem uncontroversial. To be sure, reasonable people might, for various reasons, disfavor instant replay entirely. But once we have decided to employ it, we must be trying to correct officiating errors and should therefore want (all else equal) to correct more of whatever errors there be, not fewer. Because the Pinchbeck Principle seems so unobjectionable, it will take more effort to uncover and to clearly formulate possible challenges to it.

Start with the claim that the league shouldn’t give a fig about correcting errors. As opponents of instant replay review frequently object, “an argument can be made that the human factor—in this case a blown call that costs a team dearly—is part of the game, part of its lore.”⁴⁶ Now, I’m not entirely sure what this means. That blown calls are part of the game in the sense that they have existed, do now exist, and are not entirely eliminable is beyond dispute. But as

⁴⁴ Although he confessed to giving mobsters inside information about games, for which offense he served 15 months in federal prison, Donaghy has steadfastly denied intentionally miscalling games in which he officiated.

⁴⁵ Posting of Braziled to Slate *available at* <http://fray.slate.com/discuss/forums/thread/3534640.aspx?ArticleID=2239018>

⁴⁶ Tony Long, “March 11, 1986: NFL Adopts Instant Replay,” *Wired*, Mar. 11, 2009, *available at* http://www.wired.com/science/discoveries/news/2009/03/dayintech_0311.

Hume taught, it's a fallacy to derive an "ought" from an "is". Murder, rape and robbery are similarly part of human social life. Yet surely that's no reason to tolerate more of it than we must.

The claim, accordingly, is not merely that officiating errors are ineradicable but that they are good, that we should not want to reduce them even were it costless to do so. Then-coach of the Chicago Bears Ron Rivera expressed this position emphatically after he saw his team lose to the rival Packers on a reversed call: "I can't wait for them to get rid of instant replay. . . . They have definitely taken out human error and the human nature of football. It's out. We might as well just put robots in the football game and let them play."⁴⁷ Or as George Will put it, "Human error is not a blemish to be expunged from sports, it is part of the drama."⁴⁸ True enough: sport would lose its value in a world of human infallibility. A home run is grand only in a context in which the batter might have failed. But the use to which this observation is put by defenders of a heightened standard of review—and by those who oppose instant replay entirely—relies upon a failure to distinguish human errors by players from human errors by officials. Sports law professor Howard Wasserman makes clear, when explaining his opposition to instant replay, that he equates the two. "[I]f we accept inevitable mistakes by players trying their best," he asks, "why not also accept inevitable mistakes by officials trying their best?"⁴⁹

But this question is only too easy to answer: We watch sports to witness great athletic feats, and the greatness of those feats is enhanced by the ever-present possibility, sometimes realized, of failure and error. However, we do not watch sports to see great *officiating* feats. We want the officials to remain in the background—which is why, while many commentators mocked Chief Justice Roberts's contention, made during his confirmation hearings, that "judges are like umpires," not one, as far as I'm aware, took issue with his observation that "nobody ever want to a ball game to see the umpire."⁵⁰ Simply put, unlike

⁴⁷ Fred Mitchell, *Picture Fuzzy to Bears*, CHICAGO TRIBUNE, Nov. 6, 1989, at C1, *quoted in* Oldfather & Fernholz, *supra* note 9.

⁴⁸ George Will, *The End of the Umpire? Foul!*, CHICAGO TRIBUNE, June 19, 2008, at C27.

⁴⁹ Posting of Howard M. Wasserman to Sports Law Blog, Howard Wasserman, "Against Instant Replay," Sports Law Blog, Sept. 10, 2007, *available at* http://sports-law.blogspot.com/2007_09_01_archive.html. In the same vein—though provoking the question "Does Wilson know what catchers *do*?"—Buffalo Bills owner Ralph Wilson objected to instant replay, reasoning that "it takes the human element out of the game, that it's like baseball doing away with the catcher and using a computer to call balls and strikes." Thomas George, *Owners Renew the Instant Replay*, N.Y. TIMES, March 3, 1990, *available at* <http://www.nytimes.com/1990/03/13/sports/owners-renew-the-instant-replay.html>.

⁵⁰ For criticism of the analogy, see Mitchell N. Berman, *Constitutional Interpretation: Non-originalism*, Philosophy Compass (forthcoming).

players' errors, officials' errors serve no useful purpose as foil to the greatness that we hope to glimpse. And there is no danger that efforts to reduce—and even, in places, to eliminate—officials' errors will threaten to eliminate players' errors too.⁵¹

In any event, the present question is not whether to have instant replay review. The answer to that question we are taking as settled. It is hard to fight the proposition that the rationale for instant replay is to correct officiating errors. (What else might it be?) A frontal assault on the Pinchbeck Principle is therefore extremely unpromising. The third set of arguments, then, seek to refine or modify the Pinchbeck Principle in such a way as to render false one or both of the additional premises (the error-minimization premise, and the not-prohibitively-costly premise) needed to generate a conclusion against IVE. One set of such arguments maintains that the goal should be to minimize total *weighted* errors, and that erroneous reversal is more costly than is affirmation of erroneous initial calls. A second set accepts that the goal is to minimize errors, but redefines errors in a way different from how we have understood it thus far, a redefinition that would make the second premise false.

1. *The IVE minimizes total weighted errors.* We have said that, on plausible assumptions, the preponderance-of-the-evidence standard minimizes total adjudicatory errors. It is the law's favored standard of proof when it deems errors in favor of each party to a lawsuit equally bad. But there is no necessary reason why both errors must be equally bad. And when they are not, minimizing

⁵¹ I anticipate that most readers will agree with this, on reflection. Those who are unpersuaded might reflect on the increased use of advanced technology not to review calls initially made by humans but to replace human decisionmakers in the first instance. For example, lets in tennis, wall touch in swimming, and points in fencing are all determined by technological means without relying upon direct human perception. Few fans, other than the most antiquarian or nostalgic, think that the introduction of such devices is a loss so long as adequately persuaded of their accuracy. *But see* Harry Collins, *The Philosophy of Umpiring and the Introduction of Decision-Aid Technology*, 37 J. Phil. Sport 135 (2010) (objecting to such devices when they offer a precision that substantially exceeds that which participants and spectators can perceive). In any event, nobody objects to such solutions on the grounds that they eliminate the opportunity for officiating error. Well, almost nobody. Oldfather and Fernholz argue that dramatic considerations have persuaded the NFL to preserve its low-tech chain system for measuring first downs despite the availability of several more precise high-tech alternatives. *See* Oldfather & Fernholz, *supra* note 9, at 74 (citing John Branch, *In High-Tech Game, Football Sticks to an Old Measure of Success*, N.Y. TIMES, Jan. 1, 2009, available at <http://www.nytimes.com/2009/01/01/sports/football/01chains.html?pagewanted=1&r=1>). I believe, however, that they overstate the importance to the NFL of "[t]he drama of close measurements." As I read it, the article that Oldfather and Fernholz cite suggests that the league's principal objections to laser-based alternatives are their high cost and the fact that they promise only trivial marginal increases in accuracy given that the officials' spot of the ball would remain highly imprecise.

total errors should not be the goal of an adjudicatory system. Instead, the system goal should be to minimize total *weighted* error.⁵²

As readers know, this is not abstract or hypothetical, but reflects the conventional justification for criminal law's heightened standard of proof. As in tort or breach of contract actions, a criminal trial can result in four basic outcomes, as represented by the following matrix.

| | | The Verdict | |
|-----------|----------------------|---------------------|---------------------|
| | | Guilty | Not guilty |
| The Facts | Defendant Guilty | True | False (negative) |
| | Defendant Not Guilty | False (positive) | True |

But in this context, most persons think that the two types of error are *not* equal. No doubt about it: acquitting a guilty defendant is a bad outcome. Not only does it leave him free to reoffend, but many people believe that punishing him is a good thing all on its own because, gosh darn it, he deserves it.⁵³ But as bad as that might be, most folks think that it is far worse to convict and punish someone who's innocent. This is a hard thing to establish to the satisfaction of those who don't see it intuitively, for degrees or amounts of badness (or of goodness) are elusive properties. We can't just point and say "see, this is how bad it is to be deprived of liberty and stigmatized undeservedly." Nonetheless, if

⁵² We might say that minimizing weighted error is always the goal (keeping other costs constant). But when the weights or costs of each type of error are equal, minimizing total errors and minimizing total weighted errors amount to the same thing. Either way, I am indulging the modestly simplifying assumption that the two types of correct decisions are equally good. Interesting recent work on the beyond a reasonable doubt standard of proof challenges that assumption for criminal trials. See Larry Laudan & Harry D. Saunders, *Re-Thinking the Criminal Standard of Proof: Seeking Consensus About the Utilities of Trial Outcomes*, 7 Int'l Comm. On Evidence 2, 1-33 (2009). But the assumption that, when it comes to reviewing on-field calls, correct affirmances and correct reversals are at least very close to equally good seems sound.

⁵³ What precisely is the "it" that he deserves—to be punished? to suffer? Something else? For thoughts on this important question that theorists of the criminal law routinely overlook see Mitchell N. Berman, *Two Kinds of Retributivism*, in R.A. DUFF & STUART GREEN, *THE PHILOSOPHICAL FOUNDATIONS OF THE CRIMINAL LAW* (OUP 2011).

we can ultimately do little more than admit to “a fundamental value judgment” that punishing the innocent is the greater harm or evil, it is one value judgment that seems to be widely shared.⁵⁴ (At least, it’s widely shared in Anglo-American culture, not necessarily at all times and places. Otto von Bismarck, first Chancellor of the German Empire, opined that “it is better that ten innocent men suffer than one guilty man escape.” And Felix Dzerzhinsky, founder of the Bolshevik secret police, upon learning of Bismarck’s view, raised him an execution: “Better to execute ten innocent men than to leave one guilty man alive.”⁵⁵)

There is only one way to ensure that the system will produce no erroneous convictions: abolish the criminal law, or (much the same) initiate no criminal prosecutions. That’s what we call a cure worse than the disease. The better alternative is to accept the inevitability that some innocent persons will be convicted and punished, while implementing procedural protections designed to minimize that outcome at acceptable cost. Such protections can operate at the various discrete stages in the system of criminal punishment, including at arrest and at formal charging. The procedural device that serves most directly to reduce the likelihood of convicting those innocent persons who aren’t weeded out at the earlier stages and end up being brought to trial is a heightened standard of proof. And that, of course, is the function of the “beyond a reasonable doubt” standard employed throughout Anglo-American criminal law. Instead of being instructed to adjudge the defendant guilty so long as persuaded that his guilt is more likely than not, factfinders are required to acquit unless persuaded of the defendant’s guilt beyond a reasonable doubt.

This is a pretty good way to reduce false positives. But it too comes at a price: Relative to the more-likely-than-not standard, it results in a greater number of false negatives—acquittals of persons who are guilty. This is probably obvious. After all, the whole point of ratcheting up the standard of proof is to ensure that some defendants who would have been convicted under the default more-likely-than-not standard will be acquitted. And some number of those acquitted folks are in fact guilty. So increasing the standard of proof produces two consequences: fewer convictions of innocent persons and more acquittals of guilty persons.

Less obvious to my first-year law students—but vastly important—is that the increase in erroneous acquittals outpaces the decrease in erroneous convictions, so that (on reasonable assumptions) the beyond-a-reasonable-doubt

⁵⁴ *In re Winship*, 397 U.S. 358, 372 (1971) (Harlan, J., concurring).

⁵⁵ See Alexander Volokh, *n Guilty Men*, 146 U. Pa. L. Rev. 173, 195 (1997). The delightful phrase “raised him an execution” is Volokh’s.

standard produces more *total* errors (false negatives and false positives combined) than does the more-likely-than-not standard. The idea is intuitive enough. For one thing, we have already said that the more-likely-than-not standard operates to minimize total errors. That wouldn't be so if departures from that standard produced just as few total errors.

The point can be grasped more easily, perhaps, if we translate these verbal standards into rough numerical equivalents. The more-likely-than-not standard is essentially a directive to find for the defendant unless one has at least .51 confidence in his guilt; the numerical equivalent to the beyond-a-reasonable-doubt standard is more controversial and probably even less realistic,⁵⁶ but most commentators usually peg it in the .90-.95 confidence range.⁵⁷ The only cases affected by a choice between the more-likely-than-not and the beyond-a-reasonable-doubt standards are those in which the factfinder's confidence in guilt lies someplace between .51 and, say, .90. (When the jury is less than .51 confident in the defendant's guilt, they should acquit under either standard; when they are more than .90 confident in his guilt, they should convict under either standard.) If, as we assumed earlier, jurors' beliefs about a defendant's guilt track reality pretty well in the aggregate, then it must be that a majority of defendants in these cases are in fact guilty. So for every innocent defendant who is acquitted under the beyond-a-reasonable-doubt standard but would have been convicted under the more-likely-than-not standard, *more than one* defendant who is acquitted under the heightened standard but would have been convicted under the lower standard is in fact guilty. In sum: the beyond-a-reasonable-doubt standard of proof predictably generates more *total* errors—erroneous acquittals and erroneous convictions combined—than would the more-likely-than-not standard.

Then why use it? Here we just repeat a point made above: the two types of error are not deemed to be equally costly; false positives are much worse than false negatives. So although the heightened standard produces a greater *number* of errors, the ordinary more-likely-than-not standard produces a greater *weight* of

⁵⁶ Psychologists are uncertain just how nuanced human confidence judgments actually are. No doubt the human confidence scale is not calibrated in hundredths: outside of games of chance and similar contexts in which we internalize frequentist probabilities, it would be bizarre to claim to be, say, .58 confident that *p*. Beyond that, however, about all that psychologists are agreed upon is that, between the poles of being "certain" that some proposition is true and being precisely in "equipoise," humans can reliably distinguish only among limited categories of subjective confidence, frequently grouping their judgments in such broad clusters as more likely than not, very likely, and almost certain. For a recent experimental study of how coarse or fine-grained are people's probability judgments, see Yanlong Sun et al., *Probabilistic Judgment on a Coarser Scale*, 9 Cognitive Sys. Res. 161 (2008).

⁵⁷ See, e.g., Daniel A. Krauss & Joel D. Lieberman, *Psychology in the Court Room*, 133 (Ashgate Publishing Ltd. 2009).

total errors. Put another way, when the two errors are of different magnitudes, then a standard that departs from more-likely-than-not is needed to minimize total *weighted* errors.

If all this is so, then we see a possible defense of the IVE standard for instant replay review: if the costs of mistakenly reversing a correct call are simply greater than those of mistakenly affirming an incorrect one. On this rationale, the indisputable standard of review yields a greater number of total errors, but less *weighted* error. Because this rationale would defend the system by analogizing, not to the clearly erroneous legal standard of review (as the first argument for IVE did), but to the beyond-a-reasonable-doubt legal standard of proof, then our friendly two-by-two matrix, suitably tweaked, might again provide illumination.

| | | Result of Review | |
|--------------|-----------|---|---|
| | | Affirm | Reverse |
| Initial Call | Correct | Correct Final Call | Mistaken Final Call (Erroneous Reversal) |
| | Incorrect | Mistaken Final Call (Erroneous Affirmance) | Correct Final Call |

As the matrix nicely displays, there are only two possible final outcomes for the final call: it can be correct or incorrect. But these two possibilities correspond to four permutations, for there are two different routes to each of the two possible end states. Moving clockwise from the upper left, they are: (1) a correct call affirmed; (2) a correct initial call changed to an incorrect one; (3) an incorrect initial call reversed; and (4) an incorrect initial call preserved. If a mistaken final call in the upper right (possibility 2) is worse than a mistaken final call in the bottom left (possibility 4), it must be because the value or disvalue of the final outcomes varies depending upon the route traveled to reach them: going from a correct initial call to an incorrect final call is worse than starting with an incorrect initial call and preserving it.

Why ever would that be?

Before hunting down that answer, it is time to get a clearer handle on the reason for doubt—the reason, that is, to think that the badness of a mistaken final call is independent of the route traveled to reach it.

We want the contest to proceed on the path that the players' own efforts and accomplishments warrant. If you catch the ball cleanly for a ten-yard gain, your team deserves to move ten yards down field. If you kick the ball off the crossbar and over the goal line, your team deserves to be credited with a goal. If the ball beats the batter-runner to first base, the fielding team deserves to record an out. (For extra credit: what if the ball and the batter-runner reach first at the same time?)⁵⁸ Every mistaken call—the failure to register as a reception a ball cleanly caught, the imposition of a yardage penalty for a phantom infraction, and so on—is something to be avoided and regretted precisely because it upsets this

⁵⁸ Well, as everybody knows, “the tie goes to the runner.” So if the ball and batter reach first simultaneously, you answered, the batter-runner deserves to be called “safe.”

Not so fast!

Universal sandlot wisdom about ties might trace to Major League Baseball Rule 6.05(j), which provides that “A batter is out when after a third strike or after he hits a fair ball, he or first base is tagged before he touches first base.” If the batter reaches first at the same time that first base is tagged, then the base hasn’t been tagged “before” he touches it. And so, one might reasonably surmise, he must be called safe. Unfortunately, two other rules indicate that that conclusion would be too quick.

Rule 6.09(a) provides that “the batter becomes a runner when he hits a fair ball.” That seems both obvious and unproblematic. Rule 7.08(e), however, provides that “Any runner is out when he fails to reach the next base before a fielder tags him or the base, after he has been forced to advance by reason of the batter becoming a runner.” That is, on a force play, the runner—including, per Rule 6.09(a), the batter who has become the runner—is out unless he reaches the base *before* the ball does. So it now appears that the rules are hopelessly contradictory: according to Rule 6.05(j), the tie goes to the runner; according to Rule 7.08(e), the tie goes to the fielder. Indeed, that is precisely the conclusion reached by the sleuth, the philosopher Ted Cohen, who first identified this troubling interaction among the rules in a marvelous essay. Ted Cohen, *There Are No Ties at First Base*, 79 *The Yale Review* 31 (Winter, 1990) (reprinted in Eric Bronson ed., *Baseball and Philosophy: Thinking Outside the Batter’s Box* 73-86 (2004) (explaining that when he drew this contradiction to MLB officials, advising them to remedy it, he was rebuffed on the grounds that, according to the umpires, “there never are any ties”).

But do the rules really conflict, requiring a choice between them in the case of a tie? Strictly speaking, no. Rule 6.05(j) says only that a batter-runner is out if first base is tagged before he touches it. The rule does *not* say that a batter-runner is safe if first base is tagged at the same time that he touches it. To be sure, that is the most natural implication. As Cohen puts it, that is the rule’s “import because the other rules do not give any other reason for calling him out.” I’d say, rather, that that would be the rule’s inescapable import *if* the rules did not give any other reason for calling him out. But they do—that’s the effect of Rules 6.09(a) and 7.08(e). So the apparent contradiction can be escaped by reading 6.05(j) literally, refusing to indulge its negative implication. Cf. *Royal American Oil & Gas Co. v. Szafranski*, 147 B.R. 976, 982 (Bankr. N.D. Okla. 1992) (stating the risks of negative inference as a method of statutory construction). And if we choose that path, we are left with a simple rule: ties go to the fielder!

scheme of institutional desert, of how the contest ought to proceed. *All* mistaken final calls—and only mistaken *final* calls—work this particular and most basic mischief. What makes an erroneous reversal of an initially correct call bad—something that it is to be both avoided and regretted—is precisely what makes the failure to reverse an incorrect call bad. In either case, the final call is incorrect, causing the rest of the contest to proceed on a different path than the ability and effort of the competitors themselves would have warranted. Moreover, because justice is, plausibly, the virtue of allocating goods in accordance with desert, any steps taken to reduce mistaken final calls promote justice.

To see the fundamental equivalence between erroneous reversals and erroneous affirmances more clearly—and to appreciate that there might even be something artificial or arbitrary about breaking down the officials' calls into initial call and final call—put aside instant replay review for the moment and think only about the common interplay among the on-field officials, like when the side judge initially signals a completion before the field judge rules incomplete, or when, after consultation on the field, the referee picks up a flag and announces “there was no infraction on the play.” In these cases we don't think that the initial call should be somehow protected against reversal; we want the call to be right. And when we say “the call” we mean “the final call.” Indeed, the NFL Referees' Manual is very clear on just this point. During crew conferences, it instructs referees: “Don't worry if you feel the decision making process is taking more time than you would like or looks bad. The important thing is to get it right!”⁵⁹ If this is so, then the conclusion wouldn't seem to differ even if more intermediate steps lie between initial call and final call. If we want the referee to announce the call he believes is most likely correct regardless of what the official who first weighed in believed, we should also want him to make the call he believes is most likely correct regardless of what he had first announced.

Even granting all this (as I believe we should), erroneous reversal would still be worse than erroneous affirmation *if reversal itself is a bad*. Existing arguments for IVE could be interpreted as maintaining precisely that.

(a) *Reversal is bad because it promotes disrespect for officials*. One of the most commonly voiced justifications for a highly deferential standard of review maintains that it usefully serves the league's need to protect the integrity of the game or to preserve respect for the authority of the referees. I confess that I am uncertain what “integrity of the game” means to all those who voice it.

If by “integrity” we mean something like temporal continuity, then the simple rejoinder is that football already and utterly lacks integrity of this sort,

⁵⁹ 2009 National Football League Referees' Manual, National Football League, 53 (2009).

which is one reason that Europeans, accustomed to the uninterrupted play of soccer, haven't really taken to it. Or if "integrity of the game" captures the idea that mistaken on-field calls are elements that cannot be excised without real loss then this seems much like the "blown calls are part of the lore" argument that we've just dismissed. Finally, if the phrase is meant to suggest that more frequent reversals would challenge the *moral* integrity of the officials or of other persons associated with the game, the claim should strike us as utterly implausible.

Most likely, "integrity of the game" is intended just as a fancier and more emphatic restatement of the more transparently expressed claim with which it almost invariably shares the stage: the need to preserve respect for the officials.⁶⁰ So how exactly does the IVE standard preserve respect for the officials, and just how valuable is that benefit?

The idea, I suppose, is that there is always something harmful about a *public* reversal of an on-field call *even when that reversal sets things right*. The public reversal is costly because it tends to undermine the officials' authority. To be sure, it's good if the final call is right and bad if the final call is wrong. But, on this view, it *is* worse for a final incorrect call to emerge as the reversal of an initially correct call than as the affirmance of an initially incorrect one precisely because the former contains a feature of negative value that the latter lacks: the very fact of reversal. (And, by the same reasoning, it is less good for a final *correct* call to emerge as the reversal of an initially incorrect call than as the affirmance of an initially correct one.) We can illustrate this claim by adding to our matrix the following arbitrary but illustrative values: that a call is ultimately correct confers a value of +2, that a call is ultimately incorrect confers a value of -2, and that a call has been publicly reversed confers a value of -1.

| | | Result of Review | |
|---------|---------|-------------------------|---|
| | | Affirm | Reverse (-1) |
| Initial | Correct | Correct Final Call (+2) | Mistaken Final Call (Erroneous Reversal) (-3) |
| | | | |

⁶⁰ See, e.g., Oldfather & Fernholz, *supra* note 9, at 63 ("The NFL . . . has to worry about institutional integrity. If the NFL instituted a lower standard of review for challenged plays, more calls would likely be overturned. This could eventually impact the public perception of the competency of NFL officiating crews."); Guggenheim, *supra* note 26, at 569 (opining that IVE "is crucial because it maintains the credibility of the officials and avoids fickle determinations by the referee").

| | | | |
|-------------|------------------|--|--------------------------------|
| Call | Incorrect | Mistaken Final Call (Erroneous Affirmance) (-2) | Correct Final Call (+1) |
|-------------|------------------|--|--------------------------------|

On this construal of the “respect for officials” defense of the IVE standard, the review system should be structured not to minimize the costs of error, but to minimize the costs of review, which costs are a function both of the incorrectness of final calls and the number of reversals, *inter alia*. Because the fact of reversal is itself a cost, we do in fact yield differential costs of outcome depending on the route we took to get there: a mistaken final call produced by incorrectly reversing a correct initial call *is* more costly than a mistaken final call produced by incorrectly affirming an incorrect initial call.

There may be something to this. But not nearly enough, I think, to support a standard of review as demanding as IVE.⁶¹

Most importantly, let’s not overstate the extent to which a reversal undermines respect for officials. It might well be that a widespread perception that officials are getting things wrong would threaten to breed disrespect (or less respect) for them. But it seems unlikely that such a perception will be substantially influenced by the total number of reversals of on-field calls. Fans, owners, coaches, and players get to watch and rewatch the play in super slo-mo, from a variety of camera angles, all in high definition. Therefore, we don’t need a formal announcement by the referee that an on-field call was mistaken in order to conclude that the field official got it wrong; it’s the evidence from our own eyes that licenses such a conclusion. In short, people will come to disrespect the officials if the “visual evidence” suggests to viewers at home that the officials are doing a bad job. If that’s how it appears, the fact that the calls are rarely reversed won’t protect the officials’ reputation; and if that’s not how it appears, then more frequent reversals won’t imperil it. Indeed, the fact that referees reverse their own

⁶¹ A related but distinct concern is that a lower standard would generate more reversals which would thereby threaten intra-crew harmony. Oldfather and Fernholz, *supra* note 9, at 52. This worry rings hollow to me, though I confess that I’m going mostly on untutored intuition here, and can offer no reason that even I find compelling why the reader who lacks intuitions one way or the other should credit my assessment of this worry over Oldfather and Fernholz’s contrary assessment. Still, I’ll offer two thoughts that go beyond merely reporting my sense of things. First, if reversal by the referee were a real danger to intra-crew collegiality, we might expect referees to adhere strictly to the IVE standard. If you believe, as many do, that the standard in practice is not quite as demanding as true IVE, you might take that as some evidence that intra-branch harmony is not quite so fragile. Second, insofar as there is merit to this concern, we might expect that referees operating under a formal *de novo* standard would be somewhat more deferential in practice, which prediction ought to mitigate some of the objections to relaxing the rule.

initial calls is further evidence that officials want to get calls right and are not overly sensitive about being reversed.⁶²

In any event, the IVE doesn't merely accommodate solicitude for the officials' reputation, but allows it to swamp what would seem to be a much more fundamental concern: getting the calls right. Even if some bias against reversal might be warranted, the IVE reflects a bias so pronounced as to suggest that the league has lost sight of the reason for instituting replay review in the first place.⁶³ When reputational interests and accuracy come into conflict, Major League Baseball knows where to strike the proper balance. "Umpire dignity is important," the official rule book acknowledges, "but never as important as 'being right.'"⁶⁴

Finally, not only does a concern for the officials' dignitary interests not justify the costs that IVE produces, such a high standard plausibly risks *increasing* disrespect for officials in two ways. First, insofar as the fact of reversal alone could be perceived as a rebuke, the force of that rebuke varies with the standard of review applied. Although a reversal under the IVE standard is most fairly rendered as a determination that, clearly, the initial call was mistaken, it can easily be misunderstood as a determination that the initial call was clearly mistaken. To anticipate a distinction to come, it is apt to be misperceived as a conclusion not just that the official made a mistake, but that he blundered. Reversal under the *de novo* standard does not plausibly send a similar message. In short, there will be more reversals under a *de novo* standard of review than

⁶² One of the most notorious reversals in NFL history—the call that made the “Tuck Rule” famous—was referee Walt Coleman’s overturning of his own initial call that Patriot QB Tom Brady had fumbled with less than two minutes remaining in a 2002 playoff game against the Raiders. See Richard Sandomir, *Referees Turn to Video Aid More Often*, N.Y. TIMES, Jan. 25, 2002, at D1.

⁶³ I am assuming that a nontrivial, perhaps significant, number of calls appear, after review, to be probably but not indisputably wrong. These are the calls that would come out differently under the *de novo* and IVE standards. However, some readers of this manuscript have offered the contrary speculation that, when the replays do not make it indisputably clear what the correct call would have been, then they almost invariably leave the reviewing official almost precisely in equipoise. On this view, the reviewing official will (almost) never feel pretty-confident-but-not-certain that the initial call was mistaken, in which case IVE is not merely better than *de novo*, all things considered, but virtually costless. It is hard to prove or disprove this claim without undertaking a more thorough empirical investigation. But the fact that referees so often use up the entire reviewing time allotted them is, I think, some evidence against this skeptical challenge—especially when the review results in a determination that the initial call “stands” (but is not “confirmed”). See *infra* Section V.A. As discussed earlier, it seems likely that the referee early on determines that the initial call was probably mistaken, and spends the remaining time scrutinizing the frames for confirmation.

⁶⁴ MLB Official Rules of Baseball, General Instructions to Umpires.

under IVE, but each reversal is more freighted under the IVE standard than under *de novo*.

Second, high epistemic standards impose significant psychological demands on their addressees: they require people to act other than in accordance with what they strongly believe to be true. That people will not reliably adhere to such demanding standards is suggested by the raft of experimental evidence that people are willing to convict even when significantly less confident of the defendant's guilt than the most natural readings of the "beyond a reasonable doubt" instruction would appear to permit.⁶⁵ More to the point, most observers would agree that football's own replay officials do not strictly abide by the IVE.

Take just one much-discussed example from week 15 of the 2008 season. On the road against their hated division rivals, the Baltimore Ravens, the Steelers were down 9-6 late in the fourth quarter, and driving. With 43 second left to play, and on third and goal from the Ravens' four-yard line, a scrambling Big Ben Roethlisberger hit Pittsburgh receiver Santonio Holmes camped on the front edge of the end zone. Holmes clearly had both feet in the end zone as he caught the ball with arms outstretched. But as he fell forward into the field of play, it was unclear whether the ball ever broke the plane of the end zone. Determining that it did not, the officials ruled Holmes down by contact just shy of the goal line. On review, though, referee Walt Coleman reversed the on-field call, awarding the Steelers a touchdown—and, with their 13-9 victory, the AFC North title and a first-round playoff bye.

The reversal was, to put it mildly, controversial. Indeed, just about everybody who weighed in thought Coleman got it wrong. Here's one entirely representative objection: "there's no way that last pass to Pittsburgh Steelers wide receiver Santonio Holmes on Sunday should have been called a touchdown by replay. Perhaps it could have been called a touchdown live, but not by replay. The call on the field was not an obvious mistake that needed to be corrected. It was a close call that could have gone either way, which means the call on the field should have stood."⁶⁶ Or, as *Sports Illustrated's* Peter King put it: "I didn't see the indisputable visual evidence that Walt Coleman saw under the hood in Baltimore. It was close, very close, and most likely Santonio Holmes did break

⁶⁵ Some of the experimental literature is concisely summarized in Erik Lillquist, *Recasting Reasonable Doubt: Decision Theory and the Virtues of Variability*, 36 U.C. Davis L. Rev. 85, 112-16 (2002).

⁶⁶ Patrick MacManamon, *Upon Further Review, NFL Replay System, Should Go*, AKRON BEACON-JOURNAL, available at <http://www.ohio.com/sports/mcmanamon/36179439.html>

the plane of the goal line in Baltimore on the biggest play of the biggest AFC game of the weekend, but it was not indisputable.”⁶⁷

I’m a lifelong and loyal member of Steeler Nation. So it pains me to agree with the critics. The ball might have crossed the plane in Holmes’s possession. Like most commentators, I do think it rather likelier than not. But indisputable? Not by a long shot. The critics were right.⁶⁸

As we have seen, some fans argue that *de novo* review is already the standard actually employed no matter what the rules say. That, I have said, is an overstatement. But it speaks to a widespread belief that the IVE isn’t faithfully adhered to.⁶⁹ One might reasonably surmise that these conspicuous departures from the formal standard generate disrespect for the league and its officials if anything does. To be sure, the NFL might have ways to promote greater compliance with the IVE. That’s one way to reduce the gap between what legal theorists call “the law in the books” and “the law in action.” But here’s another: lower the standard.⁷⁰

⁶⁷ Peter King, *Indisputable Evidence, Steelers continue to survive in tough games*, SPORTS ILLUSTRATED, Dec. 16, 2008, available at http://sportsillustrated.cnn.com/2008/writers/peter_king/12/14/Week15/index.html.

⁶⁸ For my own part, I’m more struck by what the complaints didn’t say than what they did. I’ve read more than a dozen, yet haven’t come across one that asks whether the rules make good sense—not one that asks why, on a play on which two teams’ playoff hopes might well rest, officials should deploy a standard that entrenches a split-second call from a side judge stationed some 30 yards away and partly screened by a Ravens defender, as against the call that appears most likely to be correct after reviewing multiple camera angles in super slo-motion. Indeed, precisely the same pattern would arise 13 months later in fans’ reactions to the pivotal calls on the Saints’ final drive against the Vikings: most agree that referee Morelli was right, given the standard of review, to let the call stand, but not one in fifty wonders why the league should employ a standard that directed Morelli to uphold a critical call that was, most fans also think, probably wrong.

⁶⁹ See Josh Patashnik, *Football, Instant Replay, and Arbitrariness*, The New Republic, Dec. 8, 2009, available at <http://www.tnr.com/blog/the-plank/football-instant-replay-and-arbitrariness> (asserting that “most referees don’t truly impose an indisputability standard”).

⁷⁰ During the 1980s, Major League Baseball confronted a gap between the law-in-the-books and the law-in-action of its own: umpires refused to call strikes as high as the rulebook required. After years of unsuccessfully trying to cajole, threaten and importune the umps to call higher strikes, MLB settled on a wise solution: it lowered the upper limit of the strike zone from the armpits to “the midpoint between the top of the shoulders and the top of the uniform pants.” Although this upper limit was still higher than what the umps had been calling, the league office reasoned that if the formal rule were closer to what the implementing officials—the umpires—liked, they’d adhere to it even if they found it less than ideal. See Murray Chass, *Baseball Changes the Strike Zone*, N.Y. TIMES, Dec. 8, 1987, available at <http://www.nytimes.com/1987/12/08/sports/baseball-changes-the-strike-zone.html>. This episode suggests a more general lesson: when the gap between the law-in-the-books and the law-in-action

(b) *Reversals are utility-reducing on net.* One theme periodically sounded in the debate over a standard of review is that proponents of *de novo* review overlook that “sports is entertainment.” One respect in which the entertainment value of sports might be thought relevant is clear enough: as we have seen, a less stringent standard of review might spell greater replay-associated delay, and delay just isn’t all that entertaining. However I believe that proponents of the objection intend it to have force independent from and additional to the objection from delay. The task is to unearth just what its import might be.

The central argument against IVE takes what might be termed a competitor’s perspective. It is not, of course, that other stakeholders are necessarily averse to correcting officiating errors: clearly, fans vary in their taste for this particular goal. It’s just that the value of error-correction gains strength from views about what those who participate in a rule-governed competition deserve, and about the justice of arranging institutions to yield outcomes that comport with desert. To charge proponents of *de novo* review with ignoring or downplaying the entertainment value of sports is, I think, to criticize them for overvaluing the interests of the participants—and even impersonal values like justice—at the expense of the desires or preferences of the fans. So why, we must ask, would the entertainment interests of fans militate against less deferential review? This is not obvious, for proponents of the entertainment-based objection to *de novo* review do not claim to have polled fans regarding which standard of review they would prefer.

One possibility, I think, is this. Roughly, spectators of any sporting event can be divided into two groups: those who care who wins that particular contest (call them “partisans”) and those who don’t (“enthusiasts”). Every call is experienced as a benefit by some of the partisans and as a detriment by others; some gain utility from the call, others lose it. So too with every reversal of a call already made: to adopt a hedonic idiom, each reversal confers pleasure on partisans of the team that benefits while inflicting pain on partisans of the team that loses. This should be uncontroversial. But its relevance might be unclear—unless it is to suggest that replay officials should generally respect a bias in favor of the home team and against the visitor, or in favor of popular teams and against the Bills and the Jaguars.

That, needless to say, is not the suggestion. The observation plausibly becomes relevant to debates over standards of review, however, if loss aversion and the endowment effect are true psychological phenomena. Briefly, the endowment effect is the experimental finding that people value goods they

becomes too great, the former can lose almost all its power to influence behavior. So, perhaps counterintuitively, a loosening of the rules might often produce a tightening of behavior.

already own more highly than an identical good they do not own. Loss aversion is the preferred explanation: people prefer avoiding losses to realizing gains.⁷¹ If these cognitive biases exist and apply in this context, then partisans want not to lose a call already made that had been to their benefit more than they want to gain the benefit of the reversal of a call already made that had been to their detriment, even when the magnitudes of the loss and the gain, in terms of probable impact on contest outcome, are the same. Similarly, if the call is reversed, it is experienced as a more profound welfare setback by partisans who lose what had been theirs than it is experienced as a welfare gain by partisans who did not “own” the call to begin with. If this is true, then holding all else constant (like the size of each team’s fan base), each reversal is utility-reducing on net. Therefore, each reversal is a cost and, as before, the total cost of an erroneous reversal is greater than the total cost of an erroneous affirmance.⁷²

All of this could be; it cannot be rejected out of hand. It would nonetheless remain to decide how to weigh this possible utility cost against the value of correcting errors. For one thing, some behavioral psychologists and economists question whether loss aversion and the endowment effect exist at all. Even if they do, the circumstances that determine whether they materialize and what their magnitudes will be are controversial.⁷³ More fundamentally, how much one thinks differential partisan utility, driven by such cognitive biases as may exist, should influence league practices at the expense of allowing more mistaken calls to go uncorrected is likely to depend in large part on the extent to

⁷¹ See, e.g., Daniel Kahneman et al., *Anomalies: The Endowment Effect, Loss Aversion, and the Status Quo Bias*, 5 J. Econ. Perspectives 193 (1991). Another cognitive bias, the omission bias, might also be relevant. But I don’t think that the omission bias adds anything to loss aversion in this particular context. And because omission bias is somewhat less well understood than is loss aversion—in particular, it is far from clear when omission bias is a genuine cognitive bias or irrationality as opposed to a moral judgment—I focus here on loss aversion only. For a recent discussion of omission bias, including a review of the literature, see Jonathan Baron & Ilana Ritov, *Omission Bias, Individual Differences, and Normality*, 94 Org. Behav. & Hum. Dec. Proc. 74 (2004).

⁷² See Patashnik, *supra* note 68.

⁷³ See, e.g., Ido Erev et al., *Loss Aversion, Diminishing Sensitivity, and the Effect of Experience on Repeated Decisions*, Max Wertheimer Minerva Center for Cognitive Studies, (2008). (empirical reevaluation of test results that were explained as proof of loss aversion and argument that these results were really a result of a diminishing sensitivity to absolute payoffs); JF Shogren et al., *Resolving Differences in the Willingness to Pay and the Willingness to Accept*, American Econ. Rev., March 1994, at 255-70 (criticizing previous tests demonstrating the endowment effect for creating a situation of artificial scarcity and providing evidence from more scientific tests that suggest no endowment effect). In light of uncertainty regarding the existence, shape, and strength of the relevant phenomena, instant replay is not just an issue that settled understandings in psychology can help address; it also provides a context in which psychological theories can be tested and refined.

which one thinks of professional football first and foremost as *sport*, thus properly governed by (among other things) norms of desert and justice, or as *entertainment*, thus properly governed by norms of happiness and preference-satisfaction.

2. *The League's goal should be to correct errors that are not (mere) mistakes.* We have already seen the claim that it's misguided even to want to eliminate officiating errors. Stated in that way, the claim is not tenable, indeed it is silly. But precisely because it is so silly, a charitable reinterpretation is in order. Perhaps the many people who express this view really mean to convey something a little different. What that might be is what I aim to explore here.

Start by marking a simple division between *beliefs* and *actions*. The officials' calls we have been discussing belong, let us say, to the world of beliefs, not of actions.⁷⁴ And we have been assuming that the belief is either right or wrong (that's the reason for excluding "judgment calls"), and that all wrong calls are "errors." In other words, to say that the goal of replay review is "to improve accuracy" or "to correct erroneous calls" is to say the very same thing. This seems to follow unexceptionally from the dictionary definition of *error* as "an act, an assertion, or a belief that unintentionally deviates from what is correct, right or true. . . . A mistake."⁷⁵

But turn attention now to the domain of actions, and consider in particular errors in baseball. An error is a misplaying that permits a batter or baserunner to reach one or more additional bases than would have been allowed given "ordinary effort" by the fielder.⁷⁶ Suppose the fielder hadn't positioned himself well, or is a little too slow to a batted ball, or doesn't put quite enough mustard on the throw. Had he not done this thing, he would have made an out. In the event, the baserunner is safe. Surely, we can't count the fielder's play a success. But we

⁷⁴ This is a crude simplification, of course, because the official doesn't merely believe or disbelieve something; he also does things, like *throws* a flag and *makes* a call. But it's not too crude for our purposes: in all but the rarest cases, the action flows directly and unproblematically from the belief—like the belief that the receiver had possession or the runner's knee was down.

⁷⁵ THE AMERICAN HERITAGE DICTIONARY OF THE ENGLISH LANGUAGE 624 (3d ed. 1992).

⁷⁶ The Official Rules are not quite as clear on this point as might be hoped. Rule 10.13 provides that "An error shall be charged for each misplay (fumble, muff or wild throw) which prologs the time at bat of a batter or which prolongs the life of a runner, or which permits a runner to advance one or more bases," but it does not explicitly define "misplay," "fumble," "muff," or "wild throw." But the notes make adequately clear that not everything that falls short of the ideal counts. Note (1) directs that "slow handling of the ball which does not involve mechanical misplay shall not be construed as an error," while Note (2) further explains that, while "it is not necessary that the fielder touch the ball to be charged with an error," the fielder's failure to do so shall be charged an error only if "in the scorer's judgment the fielder could have handled the ball with ordinary effort."

can say something stronger than that. With a little better effort or a greater exercise of skill, he could have made an out—as is not true in the cases of non-success when the fielder could not have done any better. So we can say, in cases like these, not only that the fielder’s play wasn’t a success, but that it was a failure and thus, in a sense, a mistake. Yet not all mistakes in baseball are what the sport deems an “error.”

Following the lesson of errors in baseball, we might then distinguish two types of errors: blunders and mere mistakes. So insofar as the goal of video replay in sports generally, or in football particularly, is to correct officiating errors, that statement could describe two different objectives: to correct mistake-errors or to correct blunder-errors. The argument to this point has assumed that the goal is mistake-correction. But perhaps it’s blunder-correction. If mistake-correction is the goal, then the IVE standard achieves it poorly and should be scrapped for something like *de novo* review. But perhaps mistake-correction ought not to be our ambition. Perhaps we should aim only to correct blunders-errors, not (mere) mistake-errors. If so, *de novo* review might seem—at least at first blush—to have no advantage over IVE.

Of course, it remains for defenders of IVE to explain *why* the goal should be blunder-correction and not (mere) error-correction. In truth, a satisfactory answer is not apparent to me. Nor, however, is it obvious to me that trying to develop such an argument is futile. Therefore, to stimulate further thinking on this topic, I’ll mention two paths down which a proponent of the blunder-correction construal of the Pinchbeck Principle might wish to travel. Loosely, the first path is an argument from justice, and the second an argument from virtue. Insofar as describing these argumentative strategies as “paths” suggests that they are alternatives to one another and thus not aggregable, one might prefer to think of them as building blocks thus potentially combinable. Either way, I do not claim that justice-based considerations and virtue-based ones exhaust the ways to possibly vindicate the blunder-correction principle.

The argument from justice takes issue with my earlier comments regarding the relationship between desert and justice. I have said (a) that competitors deserve what their performances warrant as measured against the rules of their sport, and (b) that justice consists of giving competitors what they deserve. A justice-based argument for blunder-correction over mistake-correction could challenge either of these claims.

Start with a challenge to proposition (a). The rules I have in mind are the rules that govern play and are addressed to the players, not the rules that govern adjudication and are addressed to the officials—the conduct rules” (or “primary

rules”), not the “decision rules” (or “secondary rules”).⁷⁷ Here’s an example that will make clearer the distinction I’m drawing. Suppose a wide receiver catches a pass in bounds for a ten-yard gain. Suppose also, however, that the officials determine, incorrectly but reasonably, that the receiver did not have possession. I have assumed that the offense deserves a first down because it has in fact moved the ball ten yards downfield and the rules governing play—the rules addressed to the competitors—provide that a ten-yard gain earns a first down. An alternative view would hold that the rules addressed to the officials must be consulted in order to determine what the participants deserve. If players deserve the outcome that accords with the instructions given the officials—instructions that are sensitive to what the officials *believe* transpired—then the offense does not deserve a first down if what the official is supposed to do *given that he believes the pass was not cleanly caught* is to rule the pass incomplete. This view strikes me as significantly less plausible than the first, but not crazy on its face.

A somewhat more promising road is to acknowledge that desert is measured by the conduct rules, but to deny that justice consists of arranging outcomes to accord with desert—it is to accept claim (a), but to challenge claim (b). It does so by exploiting a possible distinction between *desert* and *entitlement* such that the obligations of officials are taken into account not when determining what players deserve but only when determining what they are entitled to. Switching sports, suppose that the next pitch on a 3-2 count passes through the strike zone and is taken by the batter, but that the umpire mistakenly believes that the pitch was high or outside. On the view that I have favored, the pitcher *deserves* to be credited with a strikeout. Very possibly, though, he is not *entitled* to such a call. If this is right, then the important question becomes whether justice in an institutional setting consists of giving participants what they deserve or what they are entitled to. If the latter, then the idea that sports should be concerned with blunder-correction and not with mistake-correction might get off the ground if players are entitled to blunder-free officiating but not mistake-free officiating. Obviously, many dots will have to be connected to make this argument go. My aim now is only to gesture toward a possible argument not to fully sketch it out, let alone to defend it.

The second route or building block might draw from George Will’s notion that an officiating error is not a “blemish to be expunged.” Because mistake-errors are ineliminable, the argument might go, the desire to correct them reflects an unattractive instinct—perhaps even a vice of some sort. Somewhat similarly, one might worry that a reversal implies that the call being reversed fell short of what

⁷⁷ On conduct rules and decision rules, see Meir Dan-Cohen, *Decision Rules and Conduct Rules: On Acoustic Separation in Criminal Law*, 97 Harv. L. Rev. 625 (1984). On primary rules and secondary rules, see Hart, *supra* note 7, at 79-91.

was reasonably expected of the on-field official, and therefore that reversal reinforces a general social tendency for people to develop unrealistic expectations of themselves and others, which expectations undermine self-esteem and impede formation of rewarding interpersonal relationships. (A loose analogy might be drawn here to the deleterious impact that Barbie Dolls and GI Joes have on body image.) Consequently, resisting the impulse to correct mere mistakes builds character.

Frankly, I find none of this compelling. To try to correct mistakes that are correctible need not, it seems to me, either issue from or reinforce a childish and unrealistic belief in the prospects for human perfection. But others might think that these brief and telegraphic thoughts hold greater promise. Again, my goal here is only to gesture toward the sorts of ideas that, if refined and elaborated upon, might possibly serve as building blocks in support of the proposition that sports leagues should be more concerned with correcting blunder-errors and less concerned—or unconcerned—with correcting (mere) mistake errors. If that proposition can be vindicated, it is not hard to imagine the rest of the argument that would yield a conclusion in favor of highly deferential review.

D. Summary

Readers who had hoped that this Part would reach a confident conclusion regarding the single optimal standard of review for the NFL are apt now to be disappointed, for I have concluded that there remains room for reasonable disagreement regarding critical empirical and evaluative premises. In fairness, though, to have expected much more at this time was unrealistic. Before resolving a problem, you must first get a clear handle on the fruitful ways to think about it; you must identify and map out the possibly promising arguments and argumentative paths. When addressing already identified problems in mature fields, we find that much of the necessary cartographic work has already been accomplished. Not so in the case of instant replay's standard of review, a subject on which there is virtually no serious scholarly work. The same observation holds true, of course, across wide swaths of the jurisprudence of sport, which is what makes it such a promising and exciting field of scholarly inquiry.

That said, several conclusions appear warranted.

First, there is a powerful prima facie argument against IVE: (1) The reason to institute instant replay review in the first place is to correct (consequential) officiating mistakes. (2) All (final) errors are bad because they fail to award competitors the outcome they deserve in virtue of their performance. (3) Final errors produced by mistakenly reversing a correct call are not, as a class, any worse or better, *in the currency of what makes an erroneous call bad*, than final errors arrived at by mistakenly affirming an incorrect call. (4) Generally

speaking, the reviewing referee, deciding with the benefit of advanced video technology, is better able to make the correct call than are the on-field officials who have to make an initial call from a single vantage point at game speed. (5) Therefore, *de novo* review would better serve the value that instant replay is designed to serve than does a standard, like IVE, that entrenches initial calls against reversal.

Second, the most-often advanced arguments against a less deferential standard of review, like de novo, are unpersuasive: (6) Instant replay can be implemented to produce about the same amount of delay regardless of the standard of review. (7) It seems unlikely that a system of *de novo* review would diminish respect for officials any more than IVE does; indeed, the converse hypothesis seems at least as likely to be true.

Third and nonetheless, it would be premature to conclude that a non-deferential standard of review is optimal, all things considered. Perhaps the most promising route toward vindicating the IVE depends upon demonstrating that, after play has already been stopped to permit review, it is worse to reverse an initially correct on-field call than to permit to stand an initially incorrect on-field call. This would be consistent with proposition (3) above if the fact of reversal were itself a cost. Although arguments designed to show that “erroneous reversals” are worse or more costly than “erroneous affirmances” have not been well developed, we have uncovered at least one possible path toward such a conclusion: (8) Regardless of whether it corrects an officiating mistake or creates a new one, every reversal conceivably makes sports less entertaining, or reduces the pleasure and satisfaction of the spectators, because the transfer of the call reduces the utility of the losers more than it increases the utility of the winners, all else equal. Additionally or alternatively, (9) the NFL’s interest in correcting errors might be better conceived as an interest in correcting blunders than mere mistakes. While I am not myself much tempted by propositions (8) or (9), it would be premature to dismiss them. And thinking them through will require attention to, among other things: the dual character of spectator sports as both athletic contests and as popular entertainment; tradeoffs between satisfying preferences (or increasing welfare) and promoting justice; the constituents of desert and entitlement, and the relationship of each to justice and injustice; the operation of cognitive biases like loss aversion; the virtues and vices of endeavoring to correct human errors that are not also human blunders. These are rich and complex issues, of broad jurisprudential interest.

IV. THREE EASY FIXES

Given the inconclusive conclusion to Part III, the reader would be forgiven for thinking that we can say nothing more than: maybe the NFL should keep IVE, maybe it should lower the standard to *de novo*, depending on where you come down on the factors that I have agreed to be reasonably disputable. To the contrary, I will argue here that the NFL should change its system no matter what—but that different changes would be appropriate on different assumptions.

A. *Proposal 1: Correcting Mistake-Errors*

Persons who believe that the purpose of instituting instant replay should be to correct officiating mistakes at acceptable cost will be drawn to a lower standard of review that does a better job at mistake-correction. Probably the most obvious upshot of the foregoing analysis is that football's indisputable visual evidence standard of review should be scrapped and replaced with *de novo* review.⁷⁸ That's not a bad idea. But were the league to embrace *de novo* review, it could implement it in a manner that would promise even greater overall accuracy and faster review.

As the French philosopher the Marquis de Condorcet established over two hundred years ago, if individuals are even slightly more likely to answer a given question correctly than incorrectly, then increasing the number of decision makers and resolving the question by majority vote will increase the likelihood of getting the right answer. Given Condorcet's "Jury Theorem," and indulging the uncontroversial assumption that replay officials' *de novo* calls of on-field plays are more likely to be right than wrong, then resolving reviewed calls by involving some odd number of replay officials greater than one and aggregating their judgments will prove more accurate than leaving review in the hands of a single official.

So here's a concrete proposal: the NFL should set up a special video review operation in a centralized location and staff it with some modest number of trained video replay officials. Whenever a challenge arises, ad hoc panels of, say, three or five officials (or whatever number cost considerations permit) would be drawn at random from the larger pool to review all the video evidence from the

⁷⁸ To reiterate a point made earlier: *de novo* review would not render the initial call wholly unprivileged. The on-field would be "sticky" for at least two reasons. First, the burden of persuasion would rest on the challenger, so that if the reviewing official concludes that the initial call is as likely to have been right as wrong, the initial call stands. Second, the reviewing official might even take the initial on-field call as evidence in favor of the facts that would support it. When the on-field official was in a position to see the play from an angle that the cameras missed, the reviewer might reasonably infer the reason the on-field official made the call he did was because that's how things did in fact appear.

several stadium cameras much as the referee does on-site under the present system. Without deliberating, each official would register his own *de novo* judgment regarding what the call should be, and the final call would be whatever the majority “vote” established.⁷⁹ Because, in addition to exploiting Condorcet’s insight, this system would also take advantage of the greater expertise of officials specially trained in this single task, review should be more accurate as well as faster. And it would protect against a single corrupt referee even better than does the IVE.

B. Proposal 2: Correcting Blunder-Errors

The first proposal assumed that the goal of instant replay is to improve accuracy or, synonymously, to correct mistake-errors. Persons who conclude that the league should strive only to correct blunder-errors are right to disfavor *de novo* review. But here’s the interesting thing: the IVE standard *still* wouldn’t be sensible—not because it’s too deferential to the initial call, but because it’s not deferential enough! Because we can reasonably expect officials to be only as good as the visual apparatus with which nature has equipped them (subject to modest enhancement, where necessary), after-the-fact high definition, super slo-mo, multi-angle replay can disclose that the decision was indisputably substantively mistaken, even when the on-field official didn’t blunder.

To appreciate how this could be let’s take a second and even more famous balletic reception by the Steelers’ Santonio Holmes: his fully-extended grab of Roethlisberger’s high dart in the rear corner of the end zone to defeat the Arizona Cardinals in Super Bowl XLIII. As the nearly 100 million Americans who watched the contest know, although it appeared, both live and on replays, that the toes of Holmes’s right foot did touch the ground after he secured possession of the football, the video evidence remains something short of conclusive. Suppose now that NBC had used better video technology and that this improved system revealed—indisputably!—that Holmes’s toe had lifted off the turf just a fraction of an inch before he secured possession of the football.⁸⁰ In that event, we’d know that the call on the field was mistaken. But it still wouldn’t have been an *error*: Given the speed of the play, we could not reasonably have expected the official to have gotten it right; he is not fairly criticized for having gotten it wrong.

⁷⁹ Supermajority voting rules could be used instead, though this would yield more final errors than would a bare majority rule. A compromise would be to constitute reviewing panels with an even number of members and uphold the initial call if the panel is equally divided.

⁸⁰ One commentator has claimed that the video technology “used at the Beijing Olympics . . . blows away anything in use in the NFL.” Paula Duffy, *Santonio Holmes Catch Still Under Review*, *Examiner.com*, Feb. 4, 2009, available at <http://www.examiner.com/x-426-Sports-Examiner~y2009m2d4-Santonio-Holmes-catch-still-under-review>.

If the purpose of review were reconceptualized as blunder-correction not mistake-correction, it would remain to determine what standard of review should be applied to determine whether the call *was* an error in this more limited sense—a decision that an official exercising ordinary (professional) perception and judgment, having assumed his proper place on the field of play, would not have made. I have no strong views on this score. Perhaps it would be enough to instruct the reviewing referee to overturn what Little League Baseball calls “obvious wrongs”—obvious to spectators in real time, and not just obvious on review. Perhaps this is also what Australian cricket captain Ricky Ponting was thinking when observing, after instant replay was introduced to test cricket just this past year, that “we shouldn’t expect every decision to be perfect. What we’re after is to eradicate the really bad decisions.”⁸¹

Whatever the exact reviewing standard might be, the critical insight is that if our goal is to correct “really bad decisions”—what I am calling blunder-errors in contradistinction to mistake-errors—then the use of multiple-angle super slo-mo high definition video in replay might be fundamentally misguided. Think about more notorious officiating blunders: Vinnie Testaverde’s phantom touchdown against the Seahawks, or baseball umpire Phil Cuzzi’s bizarre foul call on Joe Mauer’s opposite-field fly in the 2009 American League Playoffs. What makes calls like these “really bad decisions” is that their wrongness is evident even to the naked eye. If the NFL’s goal should be, not to make sure that the call was *right*, but only that it wasn’t a blunder, review should be quicker and more perfunctory.

C. Proposal 3: That’s Entertainment

A third perspective takes very seriously the idea that partisans who lose when a call is reversed feel the sting more sharply than the partisans who gain from the reversal enjoy the windfall. Some advocates of this entertainment-oriented perspective might think the disparity is so great as to make the use of instant replay a bad buy overall. Others accept instant replay so long as the losers’ sense of loss and aggrievedness can be reduced within acceptable bounds. A highly deferential standard of review serves this interest by prohibiting reversal except in those circumstances in which even the partisans (fans and players) who lose from reversal can be expected to agree that the initial call was mistaken.⁸²

⁸¹ Graeme Smith says Test cricket referral system is half-hearted, THE AUSTRALIAN, Mar. 3, 2009, available at <http://www.theaustralian.news.com.au/story/0,25197,25132162-5001505,00.html>.

⁸² Readers who would defend the IVE might nonetheless enjoy this chestnut. Although the NFL rules rarely specify the standard of proof required for an on-field call, they sometimes do. For example, they specify a standard for receptions: if there is any doubt that the receiver made a valid reception, the pass must be ruled incomplete. NFL Rule 8, Sec. 1, Art. 6. So we can combine

That, I suspect, is the best rationale for IVE. However even if IVE ought to be preserved for this reason (or for others), three modest changes would reduce the injustice that such a demanding standard tolerates relative to a less deferential standard.

First, a team that succeeds on its challenges should get an additional challenge until it fails. Under existing rules, a team is allowed a third challenge if its first two succeed, but gets no more even if the third succeeds too. This rule is hard to justify. A challenge succeeds under IVE only if the referee has concluded that the call was *indisputably* mistaken. If a team has earned three reversals, the officials have made at least three indisputably wrong calls to that team's detriment. It seems unfair that the fact that the officiating crew had already erred so much should serve to render a team unprotected against a subsequent officiating blunder. Compare: voir dire rules do not limit a party's challenges for cause. That the defendant has already successfully challenged three members of the jury panel who were related to the plaintiff is no reason to bar her from successfully challenging the plaintiff's brother too.

In theory, this change could increase delay, though the actual impact would be trivial. A variant that probably would have a more than de minimis effect on replay-related delay would be to restore a challenge when it succeeds. That is, each coach could be granted, not two opportunities to challenge a call, but two failed challenges. This, as I say, probably would increase delays. But it wouldn't increase net delay if paired with another reform that deserves serious attention on its own: eliminate booth review in the final two minutes of each half. As Tony Dungy has recognized,⁸³ booth reviews slow the game at its most exciting juncture yet catch, and thus reverse, indisputably incorrect calls at a much lower rate than do coaches' challenges.⁸⁴ The NFL could speed up the game by placing all power to initiate review in the coaches' hands, while mitigating the impact on a coach by allowing him to challenge as many calls as he'd like until he fails twice or runs out of timeouts.

that standard of proof with the indisputable-visual-evidence standard of review to produce a nice question: if the ruling on the field is "complete," what must the replay official determine in order to overrule that call: (a) that there is indisputable visual evidence that the pass was incomplete, or (b) that there is indisputable visual evidence that it is not certain that the pass was complete? Imagine, for instance, that the camera shows only the back of the receiver as he goes out of bounds and it is impossible to tell from that angle whether he had full possession of the ball before he left the field of play, and that the official who called the pass complete was standing next to the camera, making it (arguably) indisputable that he could not have discerned with certainty that the pass was complete. What is the correct call on review?

⁸³ See John Clayton, *Dungy Wants Changes to Replay System*, ESPN, Dec. 15, 2002, available at http://a.espncdn.com/nfl/columns/clayton_john/1476343.html.

⁸⁴ See *supra* note 40.

Third and most radically, the NFL should explore the possibility of delimiting circumstances in which instant replay could be used to determine the initial call, and not just to review a call already made. The strongest case for use of replay to make the first call is on “pivotal” or “outcome-affecting” plays when the on-field officials are not confident of the correct call. Consider a pass to the end zone in the final two minutes of a close game in which, even after deliberation, the officials are uncertain whether the receiver secured possession in bounds. Sure, we’d prefer them to be confident, but they just might not be. One thing the referee can be confident of, though, is that whatever call he makes will be reviewed—either at the instigation of the replay official or (if booth reviews are eliminated) by the disappointed coach. Under such circumstances—and I do not think they are terribly rare—IVE seems to work particular unfairness. There is no good reason to forcefully entrench an initial and plainly consequential call that the on-field officials know is little more than guesswork. Rugby league and rugby union both allow match officials to defer a call on tries (scoring chances) to the replay official.⁸⁵ In short, instant replay could be used not only to review on-field calls but also—if very occasionally—to make calls in the first instance. When used in this latter manner, review must, of course, be non-deferential, and the reviewing official can be freed to reach the decision most likely to be accurate. That’s an alternative well worth considering by any system of instant replay that otherwise employs a highly deferential standard of review.

V. YES, NO, AND MAYBE: THE THREE-VERDICT SYSTEM

A. *Instant replay in the NCAA*

The NFL had experimented with video replay as early as 1986 and the system in place today dates, with revisions, back to 1999. College football, however, did not enter the replay age until the Big Ten deployed it on a trial basis in 2004. Most major conferences followed the next year and, in 2006, the NCAA enacted rules governing the use of video replay to review on-field calls. The individual conferences decide whether to use video replay but, if they choose to, they must abide by the NCAA rules.

The NCAA and NFL systems differ in detail. For example, while each coach is entitled to one challenge per game (a second if the first is successful), review in the NCAA is ordinarily initiated by a designated replay official. Also, not all the calls reviewable in the NFL are also reviewable in college football. For

⁸⁵ See *Laws of the Rugby Union 2010*, International Rugby Board, at 43, (2010). Somewhat farther afield, when a court determines that some other court (say, a higher court or the court of another jurisdiction) is better equipped to resolve a novel question of law, it is often authorized to certify that question to that other court for initial resolution.

the most part, though, the NCAA replay rules track their NFL counterparts. For purposes of this chapter, it is enough to note one fundamental similarity and one intriguing difference.

The core similarity is that the NCAA essentially parrots the NFL standard of review, directing that “to reverse an on-field ruling, the replay official must be convinced beyond all doubt by indisputable visual evidence through one or more video replays provided to the monitor.”⁸⁶ Moreover, while the NCAA has recently reiterated its commitment to this standard, it has done no better than its professional counterparts in providing a persuasive rationale.⁸⁷ In short, the NCAA sits pretty much where the NFL does: with an extraordinarily demanding standard of review and no clear articulation of the considerations or values that might support it—which is to say, no good reasons why, once video replay has been instituted, we should design it to ensure that so many wrong calls will be allowed to stand. Indeed, that the NCAA could plausibly invoke any such reasons is marginally less likely than for its professional cousin. We saw that one possible (if not especially convincing) rationale for the indisputable video evidence standard in the NFL is to reduce delay, mostly by discouraging head coaches from challenging calls in the first place. But that rationale isn’t available

⁸⁶ Rule 7, section 7. I do not know why it substitutes “video” for “visual.”

⁸⁷ In 2009 the rules committee inserted a new statement of “philosophy” to the replay rules. Codified at section 1 of Rule 12, this statement provides:

The instant replay process operates under the fundamental assumption that the ruling on the field is correct. The replay official may reverse a ruling if and only if the video evidence convinces him beyond all doubt that the ruling was incorrect. Without such indisputable video evidence, the replay official must allow the ruling to stand.

Notice to start that the second and third sentences are simply directives—indeed, they aren’t even two different directives but merely rephrasings of the very same one. So if the statement contains any defense of the standard it’s in the first sentence. No doubt the rules committee is fully warranted in assuming that the ruling on the field is correct. Without any such assumption, it might be hard to explain why the game is not stopped to permit careful review of each play as a matter of course. That would be folly. So although “the replay official and his crew shall review every play of a game,” he is sensibly instructed not to stop the game unless he believes “that there is reasonable evidence to believe an error was made in the initial on-field ruling” (and then only if he believes that “the outcome of a review would have a direct, competitive impact on the game”). But assumptions aren’t the conclusions to rigorous argument or searching inquiry. They’re starting points. As such, they are frequently mistaken—more often (we hope) right, but often enough wrong. If the video evidence leaves the replay official with a high degree of confidence—albeit short of certainty—that the on-field ruling was incorrect, then the natural thing to say is that the assumption is defeated or overcome. After all, that assumptions *can be* overcome is part of their nature as starting points for argument or action, and not as axioms or postulates. So the committee’s statement of philosophy doesn’t explain what allows us to travel from the entirely reasonable “fundamental assumption that the ruling on the field is correct” to the directive that it must be upheld when evidence strongly suggests, but does not conclusively establish, that, in this particular case, it wasn’t.

to the NCAA, where review is ordinarily initiated at the discretion of the replay official himself who may stop the game to review an unlimited number of calls and who suffers under no time limits when he does initiate a review.⁸⁸

So much for fundamental similarity. Now for the intriguing difference. The Official NFL Rules specify how certain the replay official must be before reversing an on-field call. But they don't dictate precisely what he must say. The NCAA does. If the official is sufficiently convinced (that is, convinced beyond all doubt by indisputable visual evidence) that the on-field ruling was erroneous, he must reverse the call. In doing so, he must announce what the new ruling is and specify what consequences follow. That's all straightforward enough. But what if the official is *not* convinced that the initial ruling was wrong? The rulebook recognizes two possibilities:

“1. If the video evidence confirms the on-field ruling,” then the official is directed to announce: “After further review, the ruling on the field is confirmed.”

“2. If there is no indisputable (conclusive) evidence to reverse the on-field ruling,” the official must announce: “After further view, the ruling on the field stands.”

Here, finally, is the point of interest. The two announcements—I will call them two different “verdicts”—have the same concrete consequence. Whether the replay official is persuaded that the on-field ruling was correct or merely not persuaded beyond all doubt that it was wrong, the on-field ruling stands.⁸⁹ So why bother to distinguish the two cases? Why isn't it enough just to announce either that the call stands or that it is reversed?

I don't present this question as a “puzzle.” That's because the answer seems fairly obvious. Although the first verdict—that the ruling “is confirmed”—and the second—that it (merely) “stands”—have the same tangible effects, they plainly differ in what we might term “expressive significance.” The stoppage of play and initiation of formal review broadcast doubt about the on-field official's decision. The first ruling announces to the multitude: “he got it right”; the second says only “well, I can't say that he got it wrong.” Verdict (2), to be sure, releases the referee from liability. But verdict (1) does more: it releases him from liability and *vindicates* him as well.

⁸⁸ 2009-2010 NCAA Rules and Interpretations, Rule 12, Sec. 6, Art. 2, at FR-146.

⁸⁹ An aside: the rules don't specify what it means for the video evidence to “confirm” the on-field ruling. In particular, it is unclear whether the evidence must conclusively establish that ruling's correctness, or if it is enough that the evidence leaves the replay official persuaded that the initial call was more likely than not correct.

Now, we shouldn't overdramatize. The officials sometimes get it wrong. They know it, we know it, and they know that we know it. They strive to do the best they can (and it's a tribute to their ability and training that they get the close calls right as often as they do), but they know that the best they can do is still well short of perfection. So I don't imagine that the prospect of receiving something less than explicit and unequivocal public recognition that he made the right call causes the ordinary official much anxiety—even if, in a different cultural context, things could be very different. (It might be different in sumo, for example. To this day, the ring referee, the *gyoji*, carries a dagger as a symbol of his authority. Centuries ago, the dagger served a more practical function: a *gyoji* who erred was expected to disembowel himself.)⁹⁰

Still that extra public affirmation (“he got it right”) isn't nothing—it isn't nothing for the official himself and it isn't nothing for the entity that employs and trains him. At least that's what one might reasonably have surmised in the absence of evidence one way or another. However, the very existence of the NCAA three-verdict system—(1) confirmed, (2) stands, and (3) reversed—is powerful supporting evidence. That the NCAA bothers to instruct its replay officials to distinguish between two judgments even when, in one sense, the difference doesn't matter, tells us that, in another sense, it must.

Turns out that it matters in the NFL too. Although the official rulebook itself doesn't distinguish between the two types of affirmance, actual practice on the pro level has in recent years started to track the collegiate example. Sometimes when affirming the on-field call, the referee will specify in some detail just what the video reveals. Other times, he will favor the audience only with the terse “after further review, the ruling on the field stands.” It *stands*, but it isn't *confirmed*—a distinction that several broadcast announcers have increasingly emphasized.⁹¹

B. From the Gridiron to the Courtroom

Much of this essay has been designed to show that tools and analysis developed in the law can be of help to sports leagues by helping them to more sensitively evaluate the costs and benefits of different systems of instant replay review. Now I want to argue that football can return the favor. For a moment's

⁹⁰ JAPAN: TRUE STORIES OF LIFE ON THE ROAD, 39 (Donald George & Amy Carlson eds. 1999).

⁹¹ In 2009, the league office revised its official referees' manual to validate this practice. See 2009 National Football League Referees' Manual 52-53 (“If you confirm or reverse the ruling on the field, explain the reason why. The explanation should be short and to the point. The only time you should announce ‘the ruling on the field stands’ is when there is not enough visual evidence to make a decision. However, do not say the play or ruling is ‘confirmed’ or ‘reversed’.”).

reflection should reveal that a system of three rather than two verdicts has no particular or unique linkage to instant replay or to football or to sports more generally. To the contrary, a three-verdict system is at least an option anytime a decision procedure employs a standard of proof (or of review) more demanding than the default more-likely-than-not standard.⁹² And the longest standing, most widespread, and most familiar such decision procedure is found, not in football, but in law. It is criminal law's "beyond a reasonable doubt" standard of proof.

As we have seen, that standard requires that if the jury is persuaded of the defendant's guilt beyond a reasonable doubt, they must convict. If the jury is not so persuaded, they must acquit, and the state may not retry the defendant on the same charges. In the latter case, the formal verdict is "not guilty." But that's a famous misnomer. It does not mean "We believe that the defendant *is* not guilty" but rather "We believe that the state has *failed to prove* the defendant's guilt beyond a reasonable doubt."

Here, then, is the question: Why just two verdicts? Why not three? If we think it's desirable to distinguish the two different bases for affirming an official's call in football, why isn't it also desirable to distinguish the two different bases for acquitting a defendant in a criminal trial? The state could permit (or require) the jury to select between two different forms of acquittal: "not guilty" (or "innocent") when they are persuaded that the defendant is not, in fact, guilty; and "not proven guilty" when they believe that the defendant is guilty more likely than not, but are left with too much doubt to convict. Why shouldn't we?

Many readers, I suspect, will have the immediate intuitive reaction that the football case and the criminal case are really quite different, that although a three-verdict review system makes good sense when instant replay is involved, we are better off with the existing two-verdict system in criminal law. That could be the right answer. But we'd do well not to invest too much confidence in that initial intuition or reaction, even if strongly felt. One of the most stable and robust findings in social science concerns what psychologists and economists call the status quo bias: a cognitive bias that disposes us to prefer the way things are, and to assume more strongly than the evidence would justify that the way things are is supported by sound reason. So you might be drawn to the two-verdict system because there are good reasons for it, or you might be drawn to it because that's the way things are.⁹³

⁹² Strictly speaking, a three-verdict system is possible even under the preponderance standard, for a separate verdict could apply when the factfinder is precisely in equipoise. But this is trivial: by this reasoning, football's two separate standards of proof could support *five* verdicts.

⁹³ Plausibly, the status quo bias could also explain some or much of the resistance to a lower standard of review for instant replay in football.

Well, that's not the way things are everywhere. Not in Scotland, for example. Although part of the United Kingdom, Scots law is its own legal system, tracing back to Roman law and today containing elements of English common law and the continental civil law tradition. Among the more distinctive features of Scottish criminal law is the availability of three verdicts. As in English criminal law, Scottish judges and juries are required to convict a defendant if persuaded of his guilt "beyond a reasonable doubt." In that case, the verdict is "guilty." But if not so persuaded, the factfinder must return either one of two verdicts: "not guilty" or "not proven." Orthodox understanding has it that the former expresses a judgment "that the accused definitely did not commit the crime with which he is charged," whereas the latter "means merely that the judge or jury has reasonable doubt as to the accused's guilt."⁹⁴ So instead of blithely accepting that American criminal law is as it ought to be, it's well worth trying carefully to identify, and evaluate, the costs and benefits of introducing a third verdict.⁹⁵

1. *Expressive winners and losers.* For ease of discussion, let's agree on some terms. Under either system, a jury would be instructed to return a verdict of "guilty"—that is, to convict—whenever persuaded beyond a reasonable doubt of the defendant's guilt. The present verdict that encompasses all acquittals, not distinguishing among them, is, as we know, "not guilty." Let's hypothesize a new acquittal of "innocent" for cases when the factfinder believes that, more likely than not, the accused is innocent of the charges;⁹⁶ and a new acquittal of "not

⁹⁴ Peter Duff, *The Not Proven Verdict: Jury Mythology and "Moral Panics,"* 1996 Jurid. Rev. 1, 6.

⁹⁵ In recent years, American legal journals have seen two proposals for reform along these lines. See Samuel Bray, Comment, *Not Proven: Introducing a Third Verdict*, 72 U. Chi. L. Rev. 1299 (2005); Andrew D. Leipold, *The Problem of the Innocent, Acquitted Defendant*, 94 Nw. U. L. Rev. 1297 (2000). And for a more radical reform proposal driven by a concern that jury verdicts should send clear and nuanced messages, see Paul H. Robinson, *STRUCTURE AND FUNCTION IN CRIMINAL LAW* (1997). You might think that the issue would have been mooted more exhaustively in the English legal literature given the greater salience to English lawyers and scholars of the Scottish experience. Not so. When I expressed my surprise at the dearth of English literature on the subject to John Gardner, the Glasgow-born Professor of Jurisprudence at Oxford, he had a ready explanation: "The quickest way to ensure that the English don't do something is to tell them that's the way it's done in Scotland."

⁹⁶ A defendant could be innocent for all sorts of reasons: because he had nothing at all to do with the crime charged, as in cases of mistaken identity; or because he committed the actions but lacked the criminal mental state, as in cases of accident or mistake; or because he had a valid defense, like self-defense, and so on. Let's put these variations aside for now. If any of these circumstances obtain, the defendant has not committed a criminal offense.

proven” when the factfinder believes it more likely that the accused is guilty, but has “reasonable doubts” that preclude conviction.⁹⁷ (See Table)

| Confidence in Guilt | 2-verdict system | 3-verdict system | Consequence |
|--|------------------|------------------|-------------|
| Guilty beyond a reasonable doubt | Guilty | Guilty | Convicted |
| More likely than not Guilty, but not beyond a reasonable doubt | Not Guilty | Not proven | Acquitted |
| More likely than not Innocent | | Innocent | Acquitted |

One might suppose at first blush that a choice between the two systems would not affect the conviction rate. So long as the standard of proof that the state must satisfy to win a conviction remains unchanged, then the same defendants destined to be convicted under a two-verdict system should be convicted under a three-verdict system, and vice versa. The only defendants who would be affected by a choice between the systems would be those who are acquitted. (These are the folks who find themselves in the shaded boxes in Table.) And it should be clear which of those defendants win and which lose. Those defendants who are awarded the new more favorable “innocent” acquittal would be winners: the verdict of “innocence” signals the jurors’ belief that the defendant was more likely than not innocent, a message that the jurors would

⁹⁷ Of course, we needn’t stop at three verdicts. We could attach a different verdict to each of several different confidence judgments: “innocent” for when the jury thinks the accused is innocent beyond a reasonable doubt; “not guilty” when the jury believes he’s innocent more likely than not; “not proven” when they think he’s probably guilty; and “guilty” when they’re persuaded of guilt beyond a reasonable doubt. Under this system, we’d still have only one verdict of conviction, but three verdicts of acquittal. This isn’t a crazy idea: insofar as we’re concerned about the expressive dimension of verdicts (which, recall, is the interest that would seem to drive the distinction between “confirmed” and “stands” in football), then we might want to distinguish those cases in which the jury believes the accused is probably not guilty from those in which it’s darn sure of his innocence. But each additional refinement carries only marginal benefits, and we might reasonably worry about introducing excessive complexity into jury deliberations. In any event, if a four-verdict system would be superior to a three-verdict system, then almost certainly a three-verdict system is superior to the present two-verdict system, which is the comparison that should most interest us to start.

have been unable to communicate in the two-verdict system. Those acquitted defendants who are adjudged “not proven” would be losers, for onlookers will know that these defendants could have been granted the more desirable “innocent” acquittal, and weren’t.

Presumably, the losers will outnumber the winners. (There will be more individuals in the “Not proven” cell than in the “Innocent” cell.) Nobody finds himself in the dock unless several actors in the system, each of whom could have done otherwise, elected to proceed: the police might not have arrested him, the prosecutor might have chosen not to charge him, a grand jury might have refused to indict. Admittedly, these aren’t all the most robust protections: defense attorneys quip that a grand jury will indict a ham sandwich. But they probably do serve to ensure that there’s a genuine case against the overwhelming majority of defendants who reach trial. And trial outcomes seem to bear this out: the conviction rate across American jurisdictions—putting plea bargains entirely to the side—hovers around 80% in state cases and 90% in federal prosecutions.⁹⁸ Given the several veto gates through which a criminal defendant must pass before reaching a jury, it seems doubtful that the majority of jurors who do vote to acquit believe, not merely that there remains reasonable doubt of the accused’s guilt, but that he is more likely than not to be innocent. If we’re just counting noses, the status quo might appear preferable.

But should we just count noses? We should if the magnitudes of the gains and losses are equal, if what each loser loses is as great as what each winner wins. In fact, though, that assumption seems highly improbable. As one scholar has wryly observed: “For an innocent suspect charged with a crime, there are only two possible outcomes: bad and really bad.”⁹⁹ The better outcome—acquittal—is still bad because of the burdens and anxiety of mounting a defense to criminal prosecution and also because of the reality that acquittals rarely erase all doubts about the defendant’s innocence that the facts of arrest and prosecution raise. You might be acquitted, but don’t expect all your coworkers, distant relatives, and former high school classmates no longer to suspect that you’re guilty. To be sure, even the more robust acquittal associated with a verdict of innocence would not eliminate all skepticism. But it should go a long way. This is an invaluable benefit to those who secure it. And not only to them: on the assumption that more information is generally better than less, because a two-tiered acquittal system

⁹⁸ See Bureau of Justice Statistics Bulletin, p.3 (April 2008, NCJ 221152) (reporting that, in the nation’s 75 most populous counties, 79% of trials in 2004 resulted in conviction). Federal prosecutors do even better than their state counterparts, securing convictions in 90% of the prosecutions they bring to trial. Compendium of Federal Justice Statistics, 2004, Table 4.2.

⁹⁹ Leipold, *supra* note 94, at 1301.

conveys more nuanced information than does the current single-tier system, society at large benefits as well.

Importantly, the very thing that renders the award of an innocent verdict so valuable to those who receive it—the fact that, under the present two-verdict system, so many observers think that acquitted defendants are probably guilty—also reduces the sting of a “not proven” verdict. Sure, defendants who get only the “not proven” acquittal in a three-verdict system are likely to be thought guilty by many. But the same is true if they had been acquitted as “not guilty” under a two-verdict system. In short, the gain to each of the (relatively fewer) winners is likely to be considerably greater than the loss to each of the (relatively more numerous) losers. Therefore, the aggregate gain might well outweigh the aggregate loss. Besides, the fact that the intermediate verdict in a three-verdict system has somewhat more sting than acquittal in a two-verdict system isn’t all bad. Most of the folks awarded a “not proven” acquittal would in fact be guilty, so perhaps fully deserve a little stigma even if we demand greater confidence to formally convict and to punish.¹⁰⁰

That last suggestion will strike some readers as harsh, and inconsistent with the presumption of innocence. Harsh, maybe. But not, I think, contrary to our vaunted presumption. The simple fact is that the presumption of innocence, though familiar and high-sounding, does not alone mean terribly much.¹⁰¹ All it does is assign the burden of proof to the prosecution. It doesn’t even say anything about what the standard of proof must be. Permitting convictions so long as the state proves the defendant’s guilt “more likely than not” is fully consistent with the presumption of innocence. To be sure, such a practice would *not* be consistent with the common law tradition, nor is it constitutionally permissible given Supreme Court precedent. But it’s not the presumption of innocence that rules this out, it’s the understanding that due process precludes conviction unless the state proves each element of the offense beyond a reasonable doubt. So I’m going to revise what I said above. What the presumption of innocence adds to the legal rule that the prosecution must prove each element of the offense beyond a reasonable doubt is almost nothing. All it does—and it may not even be needed for this—is to make clear that the evidence on which the prosecution may rely is limited to what it brings forth during the criminal trial: the fact finder may not

¹⁰⁰ Possibly this is why Senator Arlen Specter voted “not proven” when trying the impeachment of President Clinton. The Senate, he thought, should not convict unless persuaded of the President’s guilt beyond a reasonable doubt. And while he concluded that the House prosecutors had not carried their burden, Specter did not believe that the President deserved a free pass.

¹⁰¹ For an incisive discussion, see Larry Laudan, “The Presumption of Innocence: Material or Probatory?” 11 *Legal Theory* 4, 333-361 (2005).

speed the prosecution along, as it were, by treating the fact of the indictment itself as some evidence of the defendant's guilt.

The Scottish rule precludes conviction unless the prosecution proves all elements of the offense beyond a reasonable doubt. It is therefore consistent with the presumption of innocence. Now, you might say that it is inconsistent with the *spirit* of the presumption—or, better, with the spirit of the beyond-a-reasonable-doubt standard of proof—to allow adverse consequences *short of conviction* to attach unless the evidence supports a conviction. But that's not plausible. Consider that criminal defendants can be held without bail. That one fact alone demonstrates the falsity of the bromide that "it is a cornerstone of our system that defendant is considered innocent until proven guilty." Surely it wouldn't be kosher to lock up folks we consider innocent. Rather, it is a cornerstone of our system that nobody be *convicted of a crime and subject to criminal punishment* unless the state proves his guilt beyond a reasonable doubt. We can't get more from the presumption of innocence, and a three-verdict system doesn't run afoul of this essential precept.

The value of a more fully exculpatory acquittal is recognized by the State of California, which has for three decades permitted an acquitted defendant to petition the court for a declaration of factual innocence.¹⁰² Now, California's system is not well designed. For one thing, the burden imposed on the defendant—to establish that "no reasonable cause exists to believe" that he committed the charged offense—is far too demanding.¹⁰³ Furthermore, because the state does not provide counsel for pursuing a post-acquittal innocence petition, and because few most criminal defendants have the personal resources to invest on such a low-probability chance, few acquitted defendants even make the attempt.¹⁰⁴ Nonetheless, the instinct is sound: there *are* innocent acquitted defendants, and the benefit to them of being exonerated, not merely exculpated, is immeasurable.

To say that it's immeasurable does not mean, moreover, that it is entirely intangible. The police maintain arrest records even when the arrest results in the dismissal of charges or an acquittal at trial. So a defendant acquitted by the customary "not guilty" verdict is still more likely to be suspected and arrested for a new offense that strikes the police as similar to the old one. And an arrest can be grounds in some states for revocation of probation or parole even when it results in acquittal so long as the officer at the revocation hearing believes that the

¹⁰² Cal. Penal Code § 851.8.

¹⁰³ This is Leipold's judgment too. See Leipold, *supra* note 94, at 1324-25.

¹⁰⁴ *Id.* at 1325 & n.97.

probationer or parolee was probably guilty. Presumably, defendants acquitted as “innocent” would be relieved of these burdens.¹⁰⁵

2. *Total convictions.* So far we have supposed that the choice between the two systems would have no effect on the conviction rate: a defendant destined for conviction under the existing two-verdict system would be convicted under a three-verdict system, and vice versa. But that is almost certainly incorrect. Some number of defendants who are adjudged “guilty” under the two-verdict system would be adjudged “not proven” under a three-verdict system.

This is admittedly counterintuitive. But the claim rests on a well established psychological bias known as “the framing effect”—the phenomenon in which two “logically equivalent (but not *transparently* equivalent) statements of a problem lead decision makers to choose different options.”¹⁰⁶ In a standard example, physicians and public health officials evaluate some policy option—say, a quarantine or compulsory vaccination—differently depending on whether they are told the percentage of lives that the measure will save or the percentage of expected deaths that it will fail to avoid, even when these amount to the same thing.

One subtype of the framing effect is “the compromise effect,” the tendency of people to evaluate a given option more favorably when it appears as intermediate in a choice set rather than as an extreme. In one particularly relevant set of experiments,¹⁰⁷ a summary of facts concerning a homicide was presented to two groups of subjects. Subjects in the first group were asked to decide whether the defendant’s actions satisfied the legal definition of manslaughter (punishable by eight years imprisonment) or of murder (punishable up to life imprisonment with the possibility of parole). Subjects in the second group were offered, in addition, a third possibility of aggravated murder (punishable by death or life imprisonment without possibility of parole).

Forty seven percent of the subjects in the first group (the two-option group) chose manslaughter, while 53% chose murder. But for compromise effects, we would expect that the percentage of subjects choosing manslaughter in the second group (the three-option group) would remain about the same: as a logical matter, if the defendant should be found guilty only of manslaughter when the alternative is murder, he should still be convictable only of manslaughter when an additional *more extreme* form of murder is put on the table. As it happened, though, only 19% of the subjects in the second group chose

¹⁰⁵ *Id.*, at 1330-33.

¹⁰⁶ Matthew Rabin, *Psychology and Economics*, 36 J. Econ. Lit. 11, 36 (1998).

¹⁰⁷ See Mark Kelman et al., *Context-Dependence in Legal Decision Making*, 25 J. Leg. Stud. 287 (1996).

manslaughter, the remaining 81% dividing about equally between murder and aggravated murder. This is the compromise effect—also known as extremeness aversion—in action: murder suddenly became a more attractive option just because it became outflanked. Implications for a 3-verdict option are clear: some juries that would have convicted a defendant when given a choice between two “extremes” will acquit on the grounds of “not proven” when that is the intermediate option. Three days before the Saints would meet the Colts in Super Bowl XLIV, Freddie Peacock became the 250th person whose wrongful conviction was set aside on the strength of exculpatory DNA evidence.¹⁰⁸ Possibly some number of those innocent men and women would not have been convicted in the first place had their juries been given the option of a third verdict.

3. *Scots on the rocks.* In fact, the belief that the availability of a more favorable acquittal affects the line between the less favorable acquittal and conviction was at the root of Scottish disaffection with their three-verdict system in the early 1990s when juries in three high-profile murder prosecutions returned not-proven verdicts despite what most observers deemed evidence of guilt fully adequate to support conviction—including, in one case, the defendant’s teeth marks on the victim’s body.¹⁰⁹ In the end, the reform was resisted, but without much enthusiasm. Even Lord Advocate McCluskey, who defended the existing system, acknowledged that one wouldn’t sensibly incorporate two levels of acquittal were one “starting from scratch.”¹¹⁰ But, charmingly illustrating the status quo bias, he opined that that was little reason to actually change things.

The Scots were probably right to suppose that eliminating their top-drawer acquittal and switching to a two-verdict system would result in more convictions. Yet the extent of their dissatisfaction with their existing criminal justice regime has much more to do, not with a two-tiered acquittal system per se, but with the simply nutty way they have implemented it. Judges instruct juries that there are three verdicts, and that they may not return a verdict of guilty unless persuaded of the defendant’s guilt beyond a reasonable doubt. But they don’t instruct juries regarding how to choose between a verdict of “not guilty” and one of “not proven.” This is not a mere oversight. Inexplicably, it seems *designed*. In one case, an appeals court specifically instructed trial judges *not* to explain the

¹⁰⁸ See Innocence Project, Special Report: 250 Exonerated (Feb. 4, 2010).

¹⁰⁹ See Duff, *supra* note 93; BBC Documentary.

¹¹⁰ Duff, *The Not Proven Verdict: Jury Mythology and “Moral Panics,”* at 8; Sally Broadbridge, *The “Not Proven” Verdict in Scotland*, Standard Note SN/HA/2710, House of Commons Library, 1, 8 (May 15, 2009).

difference between the two verdicts.¹¹¹ Jurors need not even be instructed that a not proven verdict is a verdict of acquittal!¹¹²

That neither judges nor jurors have much idea what the verdicts mean is evidenced by the shockingly high rate of not guilty acquittals. We have already noted the reason to expect not proven acquittals to outpace not guilty acquittals: criminal defendants are not randomly plucked from the population; in any minimally decent system of criminal justice, guilty defendants will greatly outnumber innocent ones. Remarkably, however, four-fifths of acquittals in bench trials, and two-thirds of the acquittals in jury trials, are of the “not guilty” variety.¹¹³ The Lord Advocate, then, was right: you surely wouldn’t adopt the *Scottish* three-verdict system were you starting from scratch. That doesn’t mean, though, that you shouldn’t adopt a more defensible form of two-tiered acquittals.

In any event, and putting aside the defects of the Scots’ peculiar take on a two-acquittal system, I have agreed that more defendants are likely to be acquitted even in a well-designed two-acquittal system than in a single-acquittal system. Of course, whether that would be a good thing or a bad thing is debatable. But if we really believe that it is far worse for an innocent person to be convicted than for a guilty person to go free, and if we also believe that the beyond-a-reasonable-doubt standard, when complied with, promotes the right balance between erroneous convictions and erroneous acquittals, then we should be concerned by the (unsurprising) experimental finding that jurors do not adhere to the standard. Given all this, that a three-verdict system should marginally increase the number of acquittals weighs more plausibly in its favor than against it.

C. Summary

The previous parts of this essay used concepts and analyses familiar from the legal scholar’s toolkit to destabilize widespread acceptance of the IVE standard employed in the NFL and the NCAA. This Part has turned the tables, seeking to demonstrate that practices in football provide reason to question the two-verdict (or one-acquittal) system employed throughout Anglo-American criminal law. To be sure, we haven’t said nearly enough yet to justify confidence that substantial reform of the two-verdict system would be desirable.¹¹⁴ I hope only that we have said enough to take the question seriously.

¹¹¹ Sweeney v. HM Advocate, 2002 SCCR 131.

¹¹² Renton & Brown’s Criminal Procedure (6th ed.) ¶ 18-79.9.

¹¹³ Duff, *supra* note 93, at 7.

¹¹⁴ Both Bray and Leipold note that the costs and benefits of moving to a three-verdict system can vary across different types of offenses and might be especially hard to sort out for some sexual assault offenses. See Leipold, *supra* note 94, at 1347-48; Bray, *supra* note 94, at 1317-19.

CONCLUSION

Many sports already use instant replay in limited contexts to review calls made by on-field officials. Given recent and rapid improvements in instant replay technology, and the prospect of more to come, its use is bound to grow. Of course, a sports league or regulating body contemplating the use of instant replay does not confront a binary question—to employ it or not—but must rather address a large number of questions concerning system design. One prominent question concerns the appropriate standard of review. On this issue, observers and sports officials appear to be in widespread agreement: initial calls should be strongly entrenched against reversal. The “indisputable visual evidence” (IVE) standard used in American football is widely thought to be a successful model, suitable for adaptation elsewhere.

The near-consensus in favor of highly deferential standards of review for instant replay in sports is puzzling, however, for there is obvious reason to prefer a more neutral standard. Very simply, the ostensible purpose for instituting instant replay is to correct officiating mistakes and a less deferential standard would almost certainly achieve that goal better given today’s technology. That is not enough to conclude that highly deferential standards are misguided all things considered, for other considerations might well support such standards despite their failure to minimize officiating errors. But it is very far from self-evident just what those other considerations would be, and this essay has argued that the reasons most commonly given for IVE in professional football—that a less deferential standard that would do a better job of error-correction would create excessive delay or undermine respect for the officials—are unpersuasive.

While criticizing the standard arguments for IVE, the essay has also distilled from loose and telegraphic observations in the sparse existing literature other potentially more promising routes toward an ultimately persuasive defense. In a nutshell, the most promising argument for highly deferential review might well depend upon the proposition that, once a call has been reviewed, a decision to reverse what was initially correct is worse than a decision to uphold (equivalently, a decision not to reverse) what was initially incorrect. Indeed, I suspect that an implicit belief in the truth of this proposition motivates many defenders of a highly deferential standard. As best I can tell, though, no good arguments for this proposition have yet been developed. One important contribution of this essay is to highlight this claim and to sketch a superficially plausible path toward vindicating it—a path that invokes the entertainment value of sports and likely depends upon claims that reversal and affirmation differentially impact aggregate preference-satisfaction. An alternative route toward IVE rests upon a distinction between different types of errors—what I

have called “blunder-errors” and “mistake-errors”—married to a claim that, insofar as a given sport should be interested in correcting errors, it should be interested in correcting the former and not the latter.

As these last comments indicate, it is not this essay’s aim to conclusively resolve what is the optimal standard of review for the use of instant replay in the NFL, let alone what is the optimal standard for sports generally. Serious and careful thinking about the problem is at far too early a stage to make such an ambition realistic. Rather, my goal has been to identify a largely ignored and surprisingly challenging puzzle regarding the rules and practices of sports and to start charting some paths toward its resolution. In so doing, I also hope to demonstrate that legal scholars have much to teach, and much to learn, by taking sports seriously—not as objects of regulation *by law*, but as legal systems in their own right and thus as proper subjects for comparative and jurisprudential inquiry.