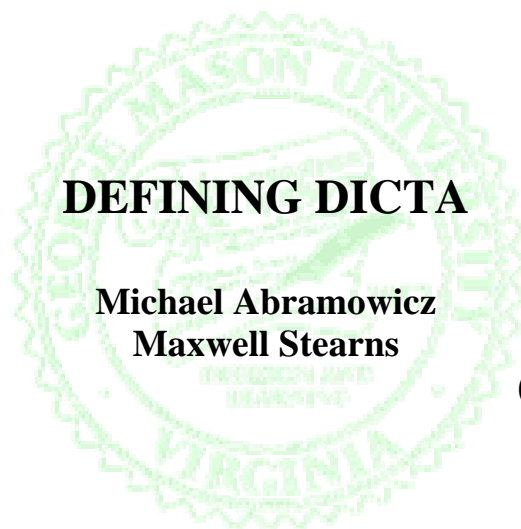


George Mason University SCHOOL of LAW



DEFINING DICTA

**Michael Abramowicz
Maxwell Stearns**

04-55

Forthcoming in
56 Stanford Law Review __ (2005)

LAW AND ECONOMICS WORKING PAPER SERIES

An electronic version of this paper can be downloaded from the following websites:
Social Science Research Network: http://ssrn.com/abstract_id=619124
bePress Legal Repository: <http://law.bepress.com/gmulwps/gmule/art14>

Defining Dicta

by Michael Abramowicz* and Maxwell Stearns**

I. INTRODUCTION	3
II. THE PROBLEMS OF DICTA	9
A. Bakke	11
1. Justice Powell’s Opinion.....	13
2. Analysis of Justice Powell’s Opinion	15
a. A Presumptive Definition	16
i. Necessary propositions and sufficient propositions	16
ii. The problem of alternative possible justifications	17
iii. The problem of alternative justifications.....	19
iv. Supportive and nonsupportive propositions	21
b. Possible Reasons for Deviating from the Presumptive Definition	22
i. The problem of structured analysis	23
ii. The problem of hypotheticals.....	29
iii. The problem of biconditional statements	31
iv. The problem of breadth	34
3. Summary and Comparison with Previous Analyses.....	35
B. Summary of Identified Problems	40
III. POSITIVE MODELS AND NORMATIVE FACTORS	42
A. Modeling Stare Decisis and the Holding-Dicta Distinction	46
1. Modeling Horizontal Stare Decisis.....	48
2. Modeling Vertical Stare Decisis	52
3. Modeling the Breadth of Holdings	59
B. The Judicial Four C’s.....	66
1. Constraint.....	67
2. Consideration.....	70
3. Clarity	72
4. Candor.....	73
C. Assessing Propositions and Problem Categories	75
1. Necessary Propositions	75
2. Sufficient Propositions and Related Problems.....	76
3. Nonsupportive Propositions.....	79
4. Structured Analysis.....	81
5. Hypotheticals	86

* Associate Professor, George Washington University Law School.

** Professor, George Mason University School of Law. For excellent comments and suggestions, we thank participants at a workshop at the George Washington University Law School. For outstanding research assistance, we thank James Boiani, David Divine, and Leah Howard. Any remaining errors should be counted as dicta.

6. Biconditional Statements	86
7. Breadth	89
8. Summary	90
IV. DEFINING HOLDING AND DICTA	92
A. The Inadequacy of Existing Definitions	93
1. Reconciliability	93
2. Factual Materiality	100
3. Necessity	103
4. Preclusivity	109
B. Necessary Conditions for a Holding	113
1. Actually Decided	116
2. Based on Facts of the Case	122
3. Leading to the Judgment	124
V. APPLICATIONS	125
A. Bakke	125
B. Roe	127
1. Justice Blackmun’s Opinion for the Court	
Error! Bookmark not defined.	
2. Applying the Definition	
Error! Bookmark not defined.	
C. Penn Central	133
D. Blakely	139
E. Grutter and Gratz	141
VI. CONCLUSION	142

I. INTRODUCTION

Even the longest, most complicated, and divisive appellate court decisions do not come with code books.¹ Appellate court decisions increasingly involve complex constitutional, statutory, or administrative law issues, and include lengthy discussions of case facts,² findings

¹ In contrast, legislatures routinely offer contemporaneous materials to assist in the task of statutory interpretation, including cross-referencing provisions and definitions of terms. *See generally* WILLIAM N. ESKRIDGE, JR. ET AL., *CASES AND MATERIALS ON LEGISLATION: STATUTES AND THE CREATION OF PUBLIC POLICY* 413, 680-82 (3d ed. 2002). Legislative history also informs statutory interpretation, and occasionally Congress will give explicit instructions about interpreting legislative history. For example, following a dispute concerning the meaning of “business necessity” for purposes of § 105 of the Civil Rights Act of 1991, the Civil Rights Act of 1991, Pub. L. 102-166, 105 Stat. 1071 (codified at 42 U.S.C. 1981 (Supp.1992)), members of Congress included a three paragraph memorandum within the Congressional Record. This was then referred to as “exclusive legislative history,” see 137 Cong. Rec. S 15,276 (daily ed. Oct. 25, 1991), and was later recognized as such within the statute. See Civil Rights Act of 1991, § 105(b). In contrast, appellate opinions do not generally include contemporaneously produced materials designed to assist future courts in discerning which aspects of those opinions are or are not controlling, and where “judicial history” exists, it is generally considered to be irrelevant to interpretation. *See generally* Adrian Vermeule, *Judicial History*, 108 *YALE L.J.* 1311 (1999) (exploring the parallels and differences between legislative and judicial history).

² For general statistics on the increasing length of judicial opinions, see RICHARD POSNER, *THE FEDERAL COURTS: CRISIS AND REFORM* 112-16 (1985).

below, hypothetical disputes of varying significance to legal issues presented, and discursive and sometimes tendentious treatment of precedent. Judges sometimes identify their holding with precision, and in so doing, imply that all other aspects of the discussion, however persuasive and seemingly relevant those discussions might be to the immediate case disposition, are instead dicta.³ More frequently, however, judges offer looser characterizations, and sometimes none at all, leaving the task of decoding dicta and holding entirely to the reader.

A judge's failure to delineate the scope of the holding within an opinion might not be a disservice to the judicial process. Even punctilious judges arguably should not be allowed the final word on the extent of their authority to resolve legal issues, and even a judge's claim to have produced a holding on a particular issue should perhaps be open to challenge when the issue seems distant from the central concerns of the case. The failure of a judicial opinion to supply reliable guidance distinguishing its holdings from its dicta, moreover, poses little difficulty to the extent that legal actors agree upon the definitions of holding and dicta. With shared understandings, future courts could be expected to follow a case's holdings and consider its dicta only to the extent that such discussions prove helpful.

Although judges and scholars share intuitions that frequently lead them to the same conclusions in particular case settings, our analysis will reveal the absence of a shared conceptual foundation for analyzing even modestly complex cases.⁴ This deficiency might reflect the tendency in recent decades of scholars interested in precedent to focus significant attention on the nature of stare decisis. A considerable literature studies the emergence, scope, and limits of stare decisis,⁵ the doctrine through which courts use opinions not merely to resolve cases, but also to make law in the form of at least presumptively binding precedents.⁶ Stare decisis plays a

³ A usage note: Like "data," "dicta" may be used as a collective noun to refer to a mass that may be singular or plural in number. See generally Dmitry N. Tychinin & Almut Beate Heinrich, *On the Correct Use of 'Data,'* 9 INT'L J. LIFE CYCLE ASSESSMENT 73 (2004), available at <http://www.scientificjournals.com/sj/lca/Pdf/aId/6510> (collecting sources on the proper use of the word "data"). While the word "dictum" may encompass only a single statement, the word "dicta" may refer to one or more statements. Thus, we use "dicta" as the analogue to "holding," which also may refer to one or more statements, and we use "dicta" whenever referring to the status of one or more propositions as holding or dicta. We use the singular "dictum" only where referring to a single proposition. Thus, we would say either "this proposition is dicta," or "this proposition is a dictum."

⁴ Judges sometimes recognize the uncertainty of the holding-dicta line, and at least one has called for more scholarly attention to the issue. See *United States v. Johnson*, 256 F.3d 895, 921 (9th Cir. 2001) (Gould, J., concurring) ("The holding/dicta debate is interesting and might be beneficial in a law journal, which could recruit academics to comment further on the issues on which my eminent colleagues disagree.").

⁵ For a discussion that draws upon a portion of that literature, see *infra* Part III.A.

⁶ See, e.g., Thomas Merrill, *Judicial Opinions as Binding Law and as Explanations for Judgments*, 15 CARDOZO L. REV. 43, 65 (1993) ("Although horizontal stare decisis creates a strong presumption that prior judicial articulations of the law are correct and should generally be followed by the rendering court, the rule is far from absolute."); Abner Mikva, *The Shifting Sands of Legal*

central role in our common law system, whether in horizontal form, for example within the Supreme Court and across federal circuit court panels, or in vertical form, for example from the Supreme Court to lower federal courts and from circuit courts to district courts.⁷ This scholarly attention is thus warranted.

As a practical matter, however, judicial analyses of precedent rarely require that courts test the contours of stare decisis doctrine directly. When stare decisis applies, a court rarely needs to consider the relatively narrow exceptions to stare decisis. Vertical stare decisis is generally considered absolute,⁸ and in the federal appellate system, en banc rehearing is required before a circuit court can overturn the precedent of a panel or an earlier en banc court. Even the Supreme Court overturns its precedents only rarely, and it debates the scope of *stare decisis* even more rarely.⁹ In contrast, evaluating a claimed precedent to determine whether an identified proposition is holding or dicta occupies a great deal of judicial attention. Indeed, before a court can decide whether to apply the doctrine of stare decisis to a given case, it must first determine

Topography, 96 HARV. L. REV. 534 (1982) (reviewing GUIDO CALABRESI, A COMMON LAW FOR THE AGE OF STATUTES (1982)) (positing that “such common law doctrines as stare decisis would presumably constrain courts applying statutes to the same extent that they constrain courts making common law decisions”); Rafael Gely, *Of Sinking and Escalating: A (Somewhat) New Look at Stare Decisis*, 60 U. PITT. L. REV. 89, 109 (2003) (observing that “common law precedents enjoy a presumption of correctness stronger than applied to constitutional cases, but not as constraining as that enjoyed by statutory precedents”). In his famous *Commentaries on American Law*, James Kent provided a similar early account of stare decisis. See 1 JAMES KENT, COMMENTARIES ON AMERICAN LAW 475 (O. W. Holmes, Jr. ed., 14th ed. Boston, Little Brown, & Co. 1896) (explaining that “[i]f a decision has been made upon solemn argument and mature deliberation, the presumption is in favor of its correctness; and the community have a right to regard it as a just declaration or exposition of the law, and to regulate their actions and contracts by it”).

⁷ At the federal level, within circuit courts, en banc panels are uniquely empowered to overturn the precedents of three-judge panels and prior en banc panels. See FED. R. APP. P. 35 (listing intracircuit conflict as one of two grounds for granting en banc review). There are, however, two caveats. First, all circuits recognize that an intervening Supreme Court decision can undermine a decision of the Court of Appeals. See, e.g., *United States v. Gay*, 967 F.2d 322, 327 (1992) (“As a general rule, one three-judge panel of this court cannot consider or overrule the decision of a prior panel. An exception to this rule arises when an intervening Supreme Court decision undermines an existing precedent of the Ninth Circuit.”). This, however, is not so much an exception to horizontal stare decisis as it is a recognition that vertical stare decisis trumps horizontal stare decisis in a lower court to which vertical stare decisis applies. See *infra* note 10. Second, some circuits follow a rule that a prior panel decision may be disregarded in “those relatively rare instances in which authority that postdates the original decision, although not directly controlling, nevertheless offers a sound reason for believing that the former panel, in light of fresh developments, would change its collective mind.” *Williams v. Ashland Engr. Co., Inc.*, 45 F.3d 588, 592 (1st Cir. 1995) (citing *Colby v. J.C. Penney Co.*, 811 F.2d 1119, 1123 (7th Cir. 1987)).

⁸ For an article that questions this conventional wisdom, see John Harrison, *The Power of Congress over the Rules of Precedent*, 50 DUKE L.J. 503, 519 (2000). While Harrison acknowledges that the Supreme Court has admonished lower federal courts not to anticipate its own overrulings, see *DeQuijas v. Shearson/American Express, Inc.* 490 U.S. 477, 484 (1989), he observes that the rule is not grounded in the Constitution and suggests that when a lower court determines that a higher court would not adhere to its own precedent, the obligation to follow vertical stare decisis might not be absolute.

⁹ For what is perhaps the most recent self-conscious exercise among Supreme Court justices considering whether to adhere to or abandon the controversial decision, *Roe v. Wade*, 410 U.S. 113 (1973), see *Planned Parenthood v. Casey*, 505 U.S. 833 (1992).

just what that case purports to establish. Because holdings in prior cases are at least presumptively binding, while dicta is not, this task requires an understanding of these terms.¹⁰

While the literature on stare decisis is broad, despite the growing need for a clear distinction to accommodate increasingly complex opinions, in recent decades the literature on the distinction between holding and dicta has been comparatively tiny.¹¹ An earlier generation of scholars, in contrast, devoted considerable attention to the holding-dicta distinction.¹² While no satisfactory definition has yet to emerge, legal scholars have largely turned their attention elsewhere. The questions whether to apply precedent, and how to construe a particular precedent in a given case, are intertwined. But they are not the same inquiry. Even an opinion without precedential value contains a holding. If anything, the more relevant inquiry in most cases is the one that has been given scant attention among the current generation of legal scholars. Courts themselves have not filled the theoretical void, and so the American judicial system lacks clearly defined rules on an important aspect of the process through which judges resolve cases and make law. Through a loose set of practices that vary considerably from jurisdiction to jurisdiction, and, perhaps more problematically, from court to court and case to case, judges define such terms as needed to assist in the task of resolving particular cases entirely on their own.

¹⁰ In our judicial system, the general rule is that courts *must* apply stare decisis to the decisions of courts above them in the judicial hierarchy. See Evan Caminker, *Why Must Inferior Courts Obey Superior Court Precedents?*, 46 STAN. L. REV. 817, 818 (1994) (observing that “longstanding doctrine dictates that a court is always bound to follow a precedent established by a court ‘superior’ to it”). This is the principle of vertical stare decisis. The force of horizontal stare decisis is not always so strong. The Supreme Court, for example, may decline to follow precedents of its own, particularly when those precedents concern constitutional law. See Mark Tushnet, *Reflections on City of Boerne v. Flores: Two Versions of Judicial Supremacy*, 39 WM. & MARY L. REV. 945 (1998) (observing that “the Supreme Court acknowledges its power to overrule its own precedents more readily in constitutional law than elsewhere”). District courts, meanwhile, may decline to follow even precedents of other district courts in the same district. See Amy Coney Barrett, *Stare Decisis and Due Process*, 74 U. COLO. L. REV. 1011, 1015 (2003) (noting that “[a]s a general rule, the district courts do not observe horizontal stare decisis”).

¹¹ Indeed, our research revealed only one major law review article in the past fifty years exclusively focused on offering a broad theoretical treatment of the distinction between holding and dicta. See Michael C. Dorf, *Dicta and Article III*, 142 U. PA. L. REV. 1997 (1994). We will find much to commend in Dorf’s analysis. See *infra* Part IV.A. Dorf’s focus, however, is on “the jurisprudential implications of Article III for determining how federal courts ought to distinguish between the holdings and dicta of past cases.” *Id.* at 2025-40, 2049-66. All courts, including courts not bound by Article III, must distinguish holdings and dicta, and our analysis will be quite general. Dorf emphasizes that courts have not applied the holding-dicta distinction consistently, *id.* at 2040-49, and explains the importance of a clear definition of the holding-dicta distinction, *id.* at 2024-28. We accept his conclusions on both points and believe that they suggest that a clearer definition of the holding-dicta distinction is necessary. Dorf himself, however, spends just a few pages on our interest, the practical problem of how to draw a distinction between holding and dicta. See *id.* at 2040-49. We will conclude that the definition Dorf offers, though inventive and arguably superior to previously offered definitions, is neither easy to apply nor normatively justifiable. See *infra* Part IV.A.4.

¹² See, e.g., EUGENE WAMBAUGH, *THE STUDY OF CASES* § 13, at 18-19 (Boston: Little, Brown & Co., 2d ed. 1894); Arthur L. Goodhart, *Determining the Ratio Decidendi of a Case*, 40 YALE L.J. 161, 180 (1930). For our analysis of the distinctions between holding and dicta offered by these scholars and others, see *infra* Part IV.A.2-3.

Despite the absence of any single governing source or universal agreement on how to define dicta, the legal system does not threaten to devolve into chaos or general incoherence. Rather, disagreements as to whether a claimed proposition is part of a court's holding, or is instead merely dicta, surface in discrete disagreements over particular cases without unraveling the fabric of the law. There is no denying, however, the importance of understanding—both as a matter of theory and at the level of practice—how to approach such a central task as sorting holding and dicta. This query goes to the heart of the business of judging, which itself goes to the essence of the Anglo-American system of interpreting and making positive law. Even if there is broad agreement on a range of issues related to decoding dicta and holdings, it should not be surprising that in the cases in which these issues matter most, the conceptual uncertainties that result from a lack of rigor in categorizing holding and dicta give rise to the greatest practical difficulties.

One difficulty in developing theoretically satisfying, and operational, understandings of the terms holding and dicta is that the most commonplace—and frequently cited—definitions of these terms are problematic in profound ways. Appreciating both why these definitions emerged and what is problematic about them is essential to our project. Consider, as perhaps the most prominent illustration, the definition of “*Obiter dictum*” in *Black’s Law Dictionary*: “[a] judicial comment made during the course of delivering a judicial opinion, but one that is unnecessary to the decision in the case and therefore not precedential.”¹³ We will argue that the definition is indefensible,¹⁴ and at least inconsistent with the general understanding that alternative holdings in a case all count as holdings.¹⁵ In fact, we will demonstrate that as a core element in the definition of holding, necessity, is itself not necessary,¹⁶ and might not even be sufficient to ensure holding status to a given proposition.¹⁷ The intuition that underlies the definition, however, is easy to appreciate, because the definition works well for simple cases. In a case in

¹³ BLACK’S LAW DICTIONARY 1102 (Bryan A. Garner ed., 8th ed. 2004); see also *infra* notes 257, 284 (distinguishing *obiter dictum* from *judicial dictum*).

¹⁴ See *infra* Part IV.A.3.

¹⁵ If a court has two possible means of achieving a disposition, for example if a challenged state law violates both equal protection and due process, then under this definition, neither basis for striking the challenged law down standing alone would be a holding. Yet, it cannot be the case that an opinion that strikes down a law on two grounds rather than one expresses no holding. For a more detailed discussion of this problem, see *infra* Part IV.A.3.

¹⁶ See *infra* Part IV.A.3 (demonstrating based upon alternative holdings that necessity is not a necessary element in affording holding status to actually decided propositions).

¹⁷ See *infra* Part IV.B.1 (suggesting that necessity might not be a sufficient condition in affording holding status to unresolved propositions that are logically required along the chosen decisional path from facts to judgment).

which there is just one issue, and just one logical argument that can take a court from the facts to the judgment, discussions that do not lie along that path are unnecessary to the decision and are therefore dicta.

In this Article, our ultimate aim is not produce a holding/dicta code book, but instead a straightforward definition of the terms “holding” and “dicta.” Like any legal test, our definition will leave some gray areas, but we are confident that the definition is both theoretically sound and functional. It reflects the issues for which we are able to achieve relatively clear resolutions, while also providing a framework for confronting those issues for which competing policy considerations render the task of providing determinate outcomes particularly difficult. We thus offer not only our recommendations for resolving even the closest conceptual issues, but also a framework with which scholars can approach such questions even if they do not embrace our specific conclusions on specific problems. By clearly identifying the issues that the literature has not directly addressed and the legal values at stake, our analysis will at least allow the courts to reach their own resolutions of these issues with greater conceptual clarity.

In the course of our analysis, we will categorize the various types of judicial assertions that can credibly be classified either as holding or dicta using a coherent and comprehensive logical structure. For example, in addition to considering the problem of alternative holdings, we will ask whether a court is generally empowered to issue what we term biconditional holdings—“if and only if” statements rather than mere “if, then” statements—one of several fundamental questions that to our knowledge has been overlooked in prior discussions of holding and dicta.¹⁸ The approach that we offer proves essential not only in ensuring analytical clarity and consistency, but also in exposing nuances that more impressionistic methodologies have failed to identify. In addition, using the methodology of rational choice, we consider the emergence of stare decisis and its role in limiting overreaching by judges in their attempts to transform dicta to holding. The analysis helps to explore both the function that stare decisis serves and how that function relates to the proper definition of holding and dicta.

We proceed as follows: Part II relies upon the famous Supreme Court case, *Board of Regents of the University of California at Davis v. Bakke*,¹⁹ to develop an analytical approach

¹⁸ The closest discussion that we have found to the biconditional statements question is Goodhart, *supra* note 12, at 180. For our analysis of Professor Goodhart’s discussion on this point and others, see *infra* Part IV.A.2 and accompanying text.

¹⁹ 438 U.S. 265 (1978).

that we use to classify various judicial assertions as either holding or dicta. Part III explores the connections between stare decisis and the distinction between holding and dicta, in tailoring appropriate judicial incentives. We begin Part III with a rational choice model that identifies endogenous incentives among judges to respect precedent, even when they disagree, and to limit the scope of opinions. Most importantly, the model exposes the limits of any resulting endogenous rules or norms, and the importance of meaningful and shared definitions of holding and dicta in enforcing and in giving content to whatever general limits on judicial overreaching that the rational choice model helps to identify. Part III then establishes the foundation for our normative critique of existing definitions of holding and dicta by evaluating the various categories of judicial assertions described in Part II based upon four normative considerations: constraint, consideration, clarity, and candor.

In Part IV, we evaluate past definitions of holding and dicta and develop our own. While our analysis in this part accommodates the analytical categories developed in Part II, and builds upon the positive and normative analyses developed in Part III, the definition of holding and dicta that we offer is at once straightforward and comprehensive. In Part IV, we propose and defend the following: “A holding consists of those propositions along the chosen decisional path or paths of reasoning that (1) are actually decided, (2) are based upon the facts of the case, and (3) lead to the judgment. If not a holding, a proposition stated in a case counts as dicta.”²⁰ While we provide a careful discussion of each element in our definition, the definition is intended to be free standing, and to provide the basis for assessing holding and dicta even in most complex appeals court decisions. In Part V, we evaluate several prominent (and complex) cases, presenting issues such as the identification of a holding in a balancing test, in light of the definition of holding and dicta that we offer.

II. THE PROBLEMS OF DICTA

We begin our inquiry with a well known Supreme Court case: *Bakke v. Board of Regents*.²¹ We are aware that this is among the most written about, and contentious, opinions in modern jurisprudence.²² Our purpose is not to take a position on or even to enter the protracted

²⁰ See *infra* Part IV.B.

²¹ 438 U.S. 265 (1978).

²² For assessments of *Bakke* and its aftermath, see JOEL DREYFUSS & CHARLES LAWRENCE, *THE BAKKE CASE* (1979); BERNARD SCHWARTZ, *BEHIND BAKKE* (1988); Lee Epstein & Jack Knight, *Piercing the Veil: William J. Brennan's Account of Regents of the*

debates as to its normative merits. To the contrary, we have chosen the case in part because more recent Supreme Court decisions on affirmative action make inquiries into what might count as holding or dicta in *Bakke* almost purely academic.²³ Nor is our purpose to focus on the unique role of the Supreme Court in developing law and public policy in such divisive matters as affirmative action, or on the relevance of particular institutional features of the Supreme Court. At least initially,²⁴ we will assume that a single definition of the distinction between holding and dicta applies to all courts, including trial and appellate courts, and that a single definition applies to both courts in which single judges hear cases and to courts in which groups of judges hear cases.

We hope that the familiarity of the case will make our exposition more accessible. We will, however, focus on features of the discussions that were not the main object of academic attention in debates on the merits of the rulings. Our aim is to use the case to clarify the analytical structure of legal reasoning by identifying the relationship between individual statements in the controlling opinion and the overall argument in that opinion. This approach will expose a set of problems about what counts as holding and what counts as dicta. *Bakke* turns out to be a particularly useful case for exploring the holding-dicta distinction because it touches on all of these problems, though at points we develop hypotheticals variations on *Bakke* to explore problems not directly implicated in *Bakke* itself. The problems that we set out vary considerably in the level of analytical difficulty that they present. When the problems that we explore have easy answers, to simplify our exposition, we will anticipate our ultimate conclusions, even though a full justification must await the analytical framework that we will develop in Part III.

Other problems, though representing reasoning patterns or structures commonly found in judicial opinions, turn out to be difficult. While our ultimate goal will be to shed light on these problems, our immediate goal is simply to identify concrete questions that previous commentators have not identified, let alone resolved, about what counts as holding and dicta. We

University of California v. Bakke, 19 YALE L. & POL'Y REV. 341 (2001); Michael Selmi, *The Life of Bakke: An Affirmative Action Retrospective*, 87 GEO. L.J. 981 (1999); and Susan Welch & John Gruhl, *Does Bakke Matter? Affirmative Action and Minority Enrollments in Medical and Law Schools*, 59 OHIO ST. L.J. 697 (1999).

²³ See *Grutter v. Bollinger*, 539 U.S. 306 (2003) (upholding the University of Michigan Law School's affirmative action program over a constitutional challenge); *Gratz v. Bollinger*, 539 U.S. 244 (2003) (striking down as unconstitutional the University of Michigan's undergraduate affirmative action program).

²⁴ We will consider the institutional difference between the Supreme Court and ordinary appellate courts, and the relevance to the definition of holding and dicta *infra* Parts III.C.3 and IV.B.

also acknowledge that in setting up our analytical framework, we adopt an admittedly technical terminology. Our ultimate objective is to provide a workable and nontechnical definition of holding and dicta that captures our eventual resolution of all the problems that we have identified.²⁵

A. Bakke

In the contentious series of cases that culminated in the Supreme Court's tandem rulings sustaining the University of Michigan Law School's affirmative action program,²⁶ and striking down that university's undergraduate affirmative action program,²⁷ a central issue for the deciding courts was to discern the meaning of *Regents of the University of California v. Bakke*.²⁸ In *Bakke*, the Supreme Court considered Allan Bakke's challenge to his rejection from the Medical School at the University of California at Davis. Bakke, who was white, alleged that the medical school denied him admission based upon race, and he argued that the medical school's use of race violated the Constitution's Equal Protection Clause.²⁹ Justice Powell issued an opinion in *Bakke* that delivered the Court's judgment.³⁰

Justice Powell concluded that the medical school's admissions program violated Bakke's equal protection rights, but he concluded that the Equal Protection Clause did not prevent the medical school altogether from considering race in its admissions process. Justice Stevens, with the support of three others,³¹ agreed with Powell's conclusion that Bakke's rights were violated. There were thus five Justices altogether who voted that Bakke be admitted. Unlike Powell, however, Stevens would have held that any use of race in admissions in a state institution of higher education violates Title VI of the Civil Rights Act of 1964.³² Meanwhile, a separate group

²⁵ See *infra* Part IV.B.

²⁶ *Grutter v. Bollinger*, 539 U.S. 306 (2003).

²⁷ *Gratz v. Bollinger*, 539 U.S. 244 (2003).

²⁸ 438 U.S. 265 (1978). *Bakke* is discussed extensively in *Grutter*, 539 U.S. at 314, 322-25, 328-30, 335, 337, 341, and in *Gratz*, 539 U.S. at 257-58, 269, 270-75.

²⁹ *Bakke*, 438 U.S. at 276-80 (opinion of Powell, J.).

³⁰ *Id.* at 269.

³¹ Chief Justice Burger and Justices Stewart and Rehnquist joined Justice Stevens. *Id.* at 408 (Stevens, J., concurring in part and dissenting in part).

³² "Race cannot be the basis of excluding anyone from participation in a federally funded program." *Id.* at 418. Justice Stevens, however, did not reach the Equal Protection Clause issue, arguing that because Bakke should prevail on statutory grounds, there was no need to consider the statutory issue. *Id.* at 412-13. Justice Stevens concurred in the judgment only "insofar as it affirms the judgment of the Supreme Court of California," adding that "[t]o the extent that it purports to do anything else, I respectfully dissent." *Id.* at 421. The words "purports to" may reinforce that Justice Stevens believed there to be an argument that some of

of four Justices, in an opinion by Justice Brennan, agreed with Powell that the medical school constitutionally could consider race,³³ and so a total of five Justices, including Justice Powell, endorsed that proposition. Unlike Powell, however, Brennan would have applied the more relaxed intermediate scrutiny test to sustain what he determined was a benign race-based classification.³⁴ Brennan would have determined that applying that test, remedying the present effects of past societal discrimination substantially furthered an important governmental interest, thus satisfying intermediate scrutiny, with the result of sustaining the Davis program.³⁵

Because of the multiple opinions and partial concurrences, the most obvious place for lower federal courts and state courts to begin their search for *Bakke*'s holding was by considering whether Powell's opinion resolved the case on the narrowest grounds, applying the framework articulated in *Marks v. United States*,³⁶ and thus counted as the controlling opinion. Indeed, this question was vigorously litigated after *Bakke*.³⁷ We believe that Powell's opinion was the controlling narrowest grounds opinion in *Bakke*,³⁸ and the Supreme Court eventually focused on the Powell opinion, albeit without conclusively resolving the *Marks* issue.³⁹ Our concern in this

Justice Powell's opinion may have been dicta.

³³ *Id.* at 326 (Brennan, J., concurring in the judgment in part and dissenting in part) (joined by Justices White, Marshall, and Blackmun).

³⁴ "[R]acial classifications designed to further remedial purposes must serve important governmental objectives and must be substantially related to achievement of those objectives." *Id.* at 359 (internal quotation marks omitted).

³⁵ *Id.* at 362-79.

³⁶ 433 U.S. 188, 193 (1977). *Marks* held that when the Supreme Court decides a case and no opinion commands majority support, the opinion consistent with the outcome that resolves the case on the narrowest grounds is controlling. *See id.*

³⁷ *See, e.g.,* Grutter v. Bollinger, 288 F.3d 732, 780-85 (6th Cir. 2002) (Boggs, J., dissenting) (arguing against the application of *Marks* to *Bakke* and providing an overview of the treatment of the *Marks-Bakke* issue in the lower courts); Johnson v. Board of Regents of University of Georgia, 263 F.3d 1234, 1248-49 (11th Cir. 2001) (refusing to apply *Marks* to *Bakke*); Hopwood v. Texas, 236 F.3d 256, 275 n.66 (5th Cir. 2000) (finding *Marks* inapplicable to *Bakke*); Smith v. University of Washington, 233 F.3d 1188, 1199-200 (9th Cir. 2000) (applying *Marks* to *Bakke*).

³⁸ We believe that *Marks* did apply to *Bakke* and that the application is neither difficult nor indeterminate. *See* MAXWELL L. STEARNS, CONSTITUTIONAL PROCESS: A SOCIAL CHOICE ANALYSIS OF SUPREME COURT DECISION MAKING 130-33 (2002) (providing detailed analysis). Justice Powell's opinion was the narrowest ground for both aspects of what he presented as the Court's judgment. Powell claimed to have resolved both that state institutions of higher learning are permitted to use race in admissions, and that the Davis approach, a quota, is not permissible. Under the narrowest ground rule, the relevant question may be cast as follows: (1) Of the opinions that would allow the use of race, which would be most restrictive in allowing race to be used? (2) Of the opinions striking down the Davis quota, which would strike down the fewest potential state law affirmative action programs? In both instances, the unmistakable answer is the opinion of Justice Powell, and his opinion is therefore the narrowest grounds opinion.

³⁹ Before the United States Supreme Court, litigants debated how to apply *Marks* to *Bakke*. *See, e.g.,* Brief for the Petitioners at 32-33, 2003 WL 164186, Gratz v. Bollinger, 539 U.S. 244 (No. 02-516); Respondent's Brief at 15 n.17, 2003 WL 402237, Gratz (No. 02-516). Once the case arrived there, however, the *Marks* analysis was beside the point. While *Marks* analysis binds lower federal courts and state courts in their efforts to construe fractured Supreme Court decisions, the Supreme Court itself is only obligated to give stare decisis effect to its prior majority decisions. *See, e.g.,* Seminole Tribe v. Florida, 517 U.S. 44, 64 (1996) (noting that only four Justices supported the reasoning of a prior precedent). Once the Supreme Court elected to reenter the affirmative action controversy, one might have expected that in the quest for *Bakke*'s meaning, all bets were off. Surprisingly, perhaps, while Justice O'Connor declined to clarify the application of *Marks* to *Bakke*, she did not maintain that doing so was

Article, however, is not with *Marks* and the narrowest grounds rule. Rather, our question is whether particular statements in an opinion count as holding or dicta, given that the opinion is controlling. We will therefore assume without further analysis that the Powell opinion is controlling, and we will place the other opinions aside. In the next section, we will provide a summary of Powell’s opinion, identifying several propositions of interest to our later analysis, and we will then analyze these propositions in the sections following.

1. Justice Powell’s Opinion

Justice Powell undoubtedly produced confusion concerning the holding in his *Bakke* opinion by making determinations concerning a program other than the one immediately before the Court. Although Powell found that the Davis medical school’s affirmative action program, which had set aside sixteen out of one hundred seats for specified racial minorities,⁴⁰ constituted an illicit quota,⁴¹ Powell also suggested that an alternative program used by Harvard University, which instead used race as one plus factor among many, and which evaluated all files as part of a combined admissions process, would withstand equal protection scrutiny.⁴² While the Regents of the University of California at Davis Medical School (the “Regents”), because they are not an elected policy making body, had improperly used race to remedy the present effects of past societal discrimination, Powell found that the Harvard program permissibly used race instead to promote its compelling interest in promoting student diversity in institutions of higher education.

To assess what counts as holding and dicta, let us break down the Powell opinion. A careful reading of Powell’s opinion reveals an attempt to resolve no fewer than the following ten propositions:

- (1) The medical school’s “special admissions program,” which set aside a prescribed number of seats for disadvantaged students,⁴³ “is undeniably a classification based on race and ethnic background.”⁴⁴
- (2) “The guarantees of the Fourteenth Amendment extend to all persons,”⁴⁵ including non-minorities, or more specifically, whites.

irrelevant to the Supreme Court’s construction of a fractured panel case. Instead, she asserted that the *Marks* analysis simply did not matter given that the *Grutter* Court agreed with the merits of Powell’s *Bakke*’s analysis. See *Grutter*, 539 U.S. at 326.

⁴⁰ *Bakke*, 438 U.S. at 275 (opinion of Powell, J.).

⁴¹ *Id.* at 315.

⁴² *Id.* at 316-18.

⁴³ *Id.* at 274-76.

⁴⁵ *Id.*

Deleted: 89.¶

(3) A race classification is permissible only when the classification satisfies the strict scrutiny test, i.e. when it is narrowly “tailored to serve a compelling governmental interest.”⁴⁶

(4) A desire to help groups of people who have suffered discrimination from society at large by admitting them on the basis of race “does not justify a classification that imposes disadvantages upon persons like [Bakke].”⁴⁷

(5) A desire to improve delivery of health care services to underserved communities was not sufficiently compelling to justify the special admissions program, because “there is virtually no evidence in the record indicating that petitioner’s special admissions program is either needed or geared to promote that goal.”⁴⁸

(6) The goal of “attainment of a diverse student body . . . clearly is a constitutionally permissible goal for an institute of higher education,”⁴⁹ and “the interest of diversity is compelling in the context of a university’s admissions program.”⁵⁰

(7) Nonetheless, by focusing on race or ethnic origin alone, the special admissions program “would hinder rather than further attainment of genuine diversity.”⁵¹

(8) A program like Harvard’s, in which “race or ethnic background may be deemed a ‘plus’ in a particular applicant’s file, yet . . . does not insulate the individual from comparison with all other candidates,”⁵² would not share this defect.

(9) Someone rejected under the Harvard plan “would have no basis to complain of unequal treatment under the Fourteenth Amendment.”⁵³

(10) “[T]hat portion of the California court’s judgment holding petitioner’s special admissions program invalid under the Fourteenth Amendment must be affirmed.”⁵⁴

We recognize, of course, that by identifying these ten propositions as representing the essence of Justice Powell’s *Bakke* opinion, we are necessarily simplifying. Justice Powell made

⁴⁶ *Id.* at 299. Justice Powell used the phrase “precisely tailored” rather than “narrowly tailored,” *id.*, but we have substituted the latter, more common statement of the strict scrutiny test.

⁴⁷ *Id.* at 310.

⁴⁸ *Id.*

⁴⁹ *Id.* at 311-12. Justice Powell stated that this conclusion had First Amendment implications. *Id.*

⁵⁰ *Id.* at 314.

⁵¹ *Id.* at 315.

⁵² *Id.* at 317.

⁵³ *Id.* at 318.

⁵⁴ *Id.* at 320.

many more statements than these, which one might parse to determine whether they are holding and dicta. We focus on these ten statements, however, because each served an important role in Powell’s analysis, and we believe that the general principles that we develop here could be applied to any remaining propositions that Powell articulated in the rest of his opinion. Even if one disputed that these ten statements accurately capture the central assertions that Powell endorsed in his analysis, doing so would not undermine our project. There will often be ambiguities about just what propositions a particular opinion endorsed, and where the boundary lines of those propositions lie. Determining the extent to which a prior opinion binds a current court requires both the resolution of ambiguities and a determination of whether propositions given such resolutions are holding or dicta.⁵⁵ Both of these are critical aspects of jurisprudential reasoning, but we are taking up only the latter challenge.

2. *Analysis of Justice Powell’s Opinion*

The following subsections will analyze Justice Powell’s opinion. We will divide our analysis into two subsections. The first subsection produces a presumptive definition of holding (and thus by negative implication, of dicta), counting all “supportive propositions” as part of the holding. A supportive proposition is one that is necessary or sufficient for the case disposition or for the disposition of another proposition that itself expresses a holding. Thus, if proposition *A* is sufficient for proposition *B*, and proposition *B* is necessary for the case disposition, then proposition *A* would count presumptively as a holding. This presumptive definition itself reflects the resolution of two significant problem categories: alternative justifications and alternative possible justifications. The second subsection considers a variety of additional problem categories, each representing a possible reason to depart from, or rebut, the presumptive definition, either by discounting a presumptive holding as dicta or by crediting presumptive dicta as holding. Constructing a definition that takes the form of a rebuttable presumption makes our analysis more manageable because it allows us to place aside, at least temporarily, some of the most difficult (and not surprisingly, more significant) issues. Our ultimate task, however, will be not only to defend our presumption-based definition, but also to offer satisfactory resolutions to the more difficult problem categories.

⁵⁵ See, e.g., *infra* text accompanying note 65 (identifying a possible ambiguity in Justice Powell’s opinion).

a. *A Presumptive Definition*

i. *Necessary propositions and sufficient propositions*

Let us start by considering the first two propositions together. Statement (1) (the Davis program is race based) might be categorized as a resolution of a mixed question of fact and law,⁵⁶ while Statement (2) (equal protection protects nonminorities, including whites) appears to be a resolution of a pure issue of law. Resolutions of both types of holdings potentially count as holdings, though holdings resolving pure questions of law may have broader application in subsequent cases.⁵⁷ Statements (1) and (2) are both *necessary propositions* to the disposition of the case in statement (10) (the program is invalid under the Fourteenth Amendment). A statement is a necessary proposition to a second statement if the court logically could not make the second statement while denying the first. On the facts of *Bakke*, the Court logically could not have concluded both that the Davis program was not race-based and that the program violated the Fourteenth Amendment. Nor could the Court have concluded that the Equal Protection Clause does not protect whites while concluding that the program violated the Clause.

The classification of statements (1) and (2) as necessary propositions to the disposition of the case could not be reached by considering statements (1) and (2) in isolation. Suppose that Bakke had also claimed that he was a victim of sex discrimination. Statement (1) then would be a necessary proposition for the proposition that the Davis plan violated Bakke's right to equal protection on the basis of race. That proposition in turn would be only a *sufficient proposition* for the conclusion that the Davis plan violated the Equal Protection Clause, because the Court could have found an Equal Protection Clause violation without finding a race-based classification. A statement is a sufficient proposition to a second statement if the court logically could not make

⁵⁶ A mixed question of law and fact asks "whether the rule of law as applied to the established facts is or is not violated." Pullman-Standard v. Swint, 456 U.S. 273, 289 n.19 (1982).

⁵⁷ Statement (1) appears to hold that any program identical to the special admissions program counts as a race classification, but it does not indicate whether other programs considering race count as racial or ethnic classifications. In theory, a subsequent court might have determined that some differently structured program did not employ a race classification even though it considered race. For example, a court might have distinguished this holding in *Bakke* in a case involving an admissions program in which there was no official policy about affirmative action, but individual admissions officers considered race. Of course, other cases might well have provided holdings that foreclosed this approach. The line between resolutions of mixed questions of law and fact and pure questions of law is often fuzzy because mixed questions can usually be restated as questions of law. See generally Evan Tsen Lee, *Principled Decision Making and the Proper Role of Federal Appellate Courts: The Mixed Questions Conflict*, 64 S. CAL. L. REV. 235 (1991) (discussing the difficulties that appellate courts face in deciding how to approach mixed questions). Justice Powell's *Bakke* opinion might have reworded statement (1) in more general terms, specifying that whenever a particular plan has certain specified features, it would be considered a race-based classification.

the first statement while denying the second. Note that in *Bakke* as written, neither statement (1) nor statement (2) is a sufficient proposition for the disposition of the case. If, for example, the court had endorsed statement (1) but rejected statement (2), thus finding that while the Davis program is race based, the Equal Protection Clause does not protect whites, the Court would not have concluded that the program was unconstitutional. As a result, statement (1) is a necessary but not a sufficient proposition for the disposition.

Crediting necessary propositions that a court has actually decided as holdings is obviously important. A legal system that did not count decided propositions that are necessary to the disposition as holdings is effectively a legal system without holdings. Thus, we can at least tentatively conclude that decided propositions that are necessary to the disposition of a case are holdings, even if those propositions are not sufficient for the result.⁵⁸ This conclusion helps explain the visceral appeal of a definition of holding as any statement necessary to resolving the case, a definition that we shall criticize in detail later.⁵⁹ But the conclusion does not justify this definition, and there is a strong argument for presumptively counting sufficient propositions to the disposition as holdings as well. The sex discrimination hypothetical may make this intuitive. Suppose that *Bakke* had raised a sex discrimination claim, and the Court concluded in one line at the end of the opinion that it did not need to consider that claim. As we have already seen, statements (1) and (2) would no longer be necessary to the result of the case (nor necessary to a proposition that itself was necessary to the outcome). Even so, one might well conclude that statements (1) and (2) should count as holdings even though the unresolved sex discrimination issue was also before the Court.

ii. *The problem of alternative possible justifications*

Statement (4) (remedying the present effects of past societal discrimination does not justify using race as a factor in admissions) and statement (5) (improving health services for underprivileged communities does not justify using race as a factor in admissions) raise the issue of sufficient propositions directly. These statements are both responses to claims by the medical school that certain social interests should save the program from being struck down. Under the

⁵⁸ For confirmation of this tentative conclusion, see *infra* Part III.C.1.

⁵⁹ See *infra* Part IV.A.3. As we have previously asserted, *see supra* notes 16-17 and accompanying text, as a free standing requirement, necessity is not necessary and might not be sufficient for a proposition to be classified as a holding. See also *infra* Parts IV.A.3, IV.B.1.

Court's selected test in Statement (3) (all race classifications are subject to strict scrutiny), which we will return to in a moment,⁶⁰ any proffered interest will suffice to save the program if and only if that interest is found to be compelling and the program is found to be narrowly tailored to further that compelling interest.⁶¹ Statement (4) appears to reject the claim that remedying the present effects of past societal discrimination is a compelling interest. Statement (5) appears to reject the claim that downstream benefits to underprivileged communities can justify an affirmative action program on narrow tailoring grounds, although the court leaves open the possibility that a plaintiff might be able to establish narrow tailoring with stronger evidence, even for a similar program.⁶² Thus, both statements (4) and (5) respectively imply sufficient propositions (lack of compelling interest, lack of narrow tailoring) for propositions (the past societal discrimination argument cannot save the statute, the health services argument cannot save the statute) that, because a contrary finding would produce the opposite result, are necessary for the disposition.

This presentation reveals our first problem category, which we will call the problem of *alternative possible justifications*. As an alternative to statement (4), Justice Powell might have concluded that the Davis plan was not narrowly tailored to remedying the present effects of past societal discrimination, but Justice Powell did not consider that issue. Similarly, as an alternative to statement (5), Justice Powell might have concluded that benefiting underprivileged communities is not a compelling interest. The issue is not whether the alternative possible justification has the status of dicta; after all, Justice Powell said nothing as to its resolution.

⁶⁰ See *infra* text following note 63.

⁶¹ *Bakke*, 438 U.S. at 299 (opinion of Powell, J.).

⁶² Statement (4) is a holding about what courts should do in the *absence* of evidence on the record about downstream benefits of diversity. On its own terms, statement (4) does not purport to determine whether a university could rely on the benefits of an affirmative action program in generating doctors for underserved communities. Thus, while *Bakke* resolved the question whether an affirmative action program could be justified by its benefits for those advantaged, it did not resolve whether it could be justified by its benefits for those not advantaged. It was at least conceivable in *Grutter* that the Court could have stated that the University of Michigan Law School, unlike the Davis Medical School, had presented data sufficient to demonstrate that the challenged affirmative action program significantly promoted access to legal services among minorities and the poor. See Brief of Amicus Curiae Michigan Black Law Alumni Society in Support of Respondents at 21-24, 2003 WL 537217, *Grutter v. Bollinger*, 539 U.S. 306 (No. 02-241) (making this argument). The basis for Justice Powell's rejection of statement (4), unlike that for his rejection of statement (3), invited the future possibility of such a fact-based distinction.

Interestingly, in *Grutter*, Justice O'Connor ignores this distinction even though it might have made her argument more persuasive. O'Connor states, "Justice Powell rejected an interest in 'increasing the number of physicians who will practice in communities currently underserved,' concluding that even if such an interest could be compelling in some circumstances the program under review was not 'geared to promote that goal.'" *Grutter*, 539 U.S. at 324 (quoting *Bakke*, 438 U.S. at 306, 310 (opinion of Powell, J.)). This statement mischaracterizes Justice Powell, who stated only that there was insufficient evidence that the program was geared to promote that goal.

Instead, it is whether the potential alternative means of reaching a conclusion through the alternative possible justification renders the actual means of reaching the conclusion dicta. Eventually, we will conclude that the answer is “no,” and that the existence of alternative possible justifications does not turn what otherwise would count as holding into dicta.⁶³

Statement (3) (all race classifications are subject to strict scrutiny) also raises the problem of alternative possible justifications, though in a more subtle way. Justice Powell’s opinion does not reach a conclusion about whether the Davis program would have failed intermediate scrutiny, had that been the appropriate test. Thus, as an alternative means of justifying the conclusion that Davis’s race classification was impermissible, Justice Powell might have found that the Davis program failed intermediate scrutiny. This conclusion would have relieved Justice Powell of the need to determine whether strict scrutiny or intermediate scrutiny was the appropriate test since a state proffer that fails intermediate scrutiny necessarily fails the more stringent strict scrutiny. Justice Powell’s opinion can be read as assuming *arguendo* that the Davis program would pass intermediate scrutiny. The problem of alternative possible justifications thus arises generally when an opinion author either explicitly or implicitly makes an assumption that, if successfully refuted, would provide further justification for the eventual result.

iii. The problem of alternative justifications

Statement (8) (a plus factor plan would not hinder the attainment of diversity) might at first appear to be neither necessary nor sufficient for any other proposition. Regardless of whether facts about Harvard’s plan were formally part of the record, Harvard’s plan was not the plan that the Court was evaluating. This analysis, however, turns out to be simplistic, because it could be argued that statement (8) and statement (7) (the Davis program hinders genuine diversity) are both sufficient propositions for the broader proposition that the Davis program fails narrow tailoring. Although Justice Powell does not explicitly identify statement (7) as responsive to the narrow tailoring analysis, a demonstration that a program hinders a goal shows *a fortiori* that the program is not narrowly tailored to achieving that goal. Identification of a more narrowly tailored program also may show that the challenged program is not narrowly tailored.

Justice Powell prefaced his comments on the Harvard plan as follows: “The experience of other university admissions programs . . . demonstrates that the assignment of a fixed number of

⁶³ See *infra* Part III.C.2.

places to a minority group is not a necessary means toward that end.”⁶⁴ Powell’s discussion of the Harvard plan might be seen as sufficient to resolve the constitutionality of the Davis plan. The discussion of the Harvard plan demonstrated that an alternative plan existed that was more narrowly tailored than the Davis plan. Thus, even if the Court had not concluded that the Davis program hindered genuine diversity, which itself was a sufficient basis for concluding that a program is not narrowly tailored, Justice Powell can be interpreted as having believed that his analysis of the Harvard plan established that the Davis plan was not narrowly tailored.

There remains, to be sure, ambiguity about whether Justice Powell actually considered the existence of a more narrowly tailored plan by itself sufficient to conclude that the Davis plan was not narrowly tailored. Depending upon how it interprets the narrow tailoring prong of strict scrutiny,⁶⁵ a future court could argue that, at best, the Harvard plan discussion contributed to Powell’s analysis only in an informal way. Often, it is possible to identify an argument that contributes to the analysis in an opinion, without being able to determine definitively whether that analysis was necessary to or sufficient for some higher-order proposition, and ultimately for the disposition of the case. Just as our task is not to resolve ambiguities concerning what propositions an opinion endorsed,⁶⁶ so too is it not to resolve ambiguities concerning what role a particular proposition played within the internal reasoning of an opinion writer, where a judge has not objectively explained how that proposition supports the judge’s argument. The existence of the ambiguity here would provide sufficient justification for a later judge to determine that statement (8) is not necessary or sufficient for any other proposition, and we will consider the statement under that assumption momentarily.⁶⁷ But for now, we will assume that statement (8) represents a sufficient proposition for the proposition that the Davis plan was not narrowly tailored.

Thus understood, statements (7) and (8) present the problem of *alternative justifications*. Justice Powell could have justified his conclusion that the Davis plan was not narrowly tailored (a conclusion that itself is not directly represented by any of the ten statements listed above) with either statement (7) or statement (8), but he chose unnecessarily to consider and endorse *both*

⁶⁴ 438 U.S. at 316 (opinion of Powell, J.).

⁶⁶ See *supra* text accompanying note 55.

⁶⁷ See *infra* text accompanying note 89.

Deleted: ¶

statements (7) and (8). The previously considered problem of alternative possible justifications presented a situation in which it might not have been necessary for an opinion author to consider one of two (or more) statements, but in which we cannot be sure, because we could not know what the analysis of the unexplored issue would have revealed. With alternative justifications, it is unmistakable that the opinion author did more than was necessary. Thus, there may be a stronger case for excluding alternative justifications from the status of holding than there would be for excluding statements representing alternative possible justifications.⁶⁸ If we were to accept the simplistic “necessary” definition of holding, alternative justifications would not be holdings.⁶⁹ Eventually, however, we will conclude that neither problem by itself should transform a statement from holding into dicta.⁷⁰

iv. Supportive and nonsupportive propositions

Statements (6) (diversity is a compelling interest) and (7) (focusing exclusively on race would hinder genuine diversity), reflect the two prongs of the Court’s strict scrutiny analysis for the diversity defense of affirmative action. In contrast, we have seen that Justice Powell used statement (4) (remediating the present effects of past societal discrimination does not justify using race as a factor in admissions) to refute the past societal discrimination defense by considering only the compelling state interest prong, and used statement (5) (improving health services for underprivileged communities does not justify using race as a factor in admissions) to refute the downstream benefits defense by considering only the narrow tailoring prong. Statement (7) is like statements (4) and (5), however, in the sense that all three statements articulate propositions that are sufficient to support a proposition that is necessary to the eventual result. Specifically, by rejecting narrow tailoring, statement (7) implies the proposition that on the facts of the case, the state’s proffered goal of promoting academic diversity will not save the program. Thus,

⁶⁸ When a judge advances multiple justifications, there is a danger that the judge is doing so because the judge wishes to have control over the resolution of multiple issues. See *infra* Part III.B.1, Part III.C.2. One might also make other arguments for counting alternative justifications as holding, while ignoring alternative possible justifications. For example, one might argue that alternative justifications invite an intellectual shell game in which a judge who perceives weaknesses in argument *A* advances argument *B*, and perceiving weaknesses in argument *B* advances argument *C*. But because none of the arguments needs to be independently conclusive—instead they are each alternative justifications—the judge does not iron out the wrinkles in any of the arguments (or reject any or all of the arguments as deficient). This problem is less acute in the context of alternative possible justifications because there the court has chosen to resolve the issue on one ground, but has not foreclosed the possibility that a different court selecting a different ground might get to the same judgment.

⁶⁹ See *infra* note x and accompanying text.

⁷⁰ See *infra* Part III.C.2.

statement (7) (a quota undermines diversity) is sufficient to a proposition (the Davis program is not narrowly tailored) that is necessary for the ultimate result (striking the program).

Unlike statements (4) and (5), however, statement (7) does not present the problem of alternative possible justifications. Justice Powell explored the apparent alternative justification (the compelling interest prong) in statement (6) (diversity is a compelling interest), thus rejecting the argument that diversity provides no compelling interest as a possible alternative justification for his result. Thus, statement (7) would appear to have at least as much of a claim to holding status as statements (4) and (5). Statement (6), in contrast, would appear to have less of a claim to the status of holding. At least in our stylized presentation, it appears that it would have made no difference to the case outcome, or to his need to consider other legal issues, had Justice Powell skipped statement (6).

More concretely, statement (6) is not a “supportive proposition,” by which we mean one that is necessary to or sufficient for either the disposition or any other supportive proposition. Identifying “nonsupportive propositions” like statement (6) is useful because such propositions can be freely omitted without affecting either the result or the need for a court to determine other propositions en route to reaching the result. Returning to our earlier hypothetical, if the Court had concluded that the Davis plan also constituted sex discrimination against Bakke, that would be a supportive proposition. That determination would not have affected the result, because absent the discussion the Court would have achieved the same judgment relying solely on race discrimination. The determination, however, would have meant that the propositions related to race discrimination could alternatively have been omitted without affecting the result, since the Court could have relied solely on sex discrimination. But if the Court had concluded that the Davis plan did not constitute sex discrimination against Bakke, that would be a nonsupportive proposition, because inclusion of such reasoning would have had no effect on the remainder of the analysis.

b. Possible Reasons for Deviating from the Presumptive Definition

We will assume for now that a supportive proposition *presumptively* counts as a holding. As a result, any proposition that is necessary to or sufficient for either the disposition or for another holding will itself be a presumptive holding. This assumption will allow us to separate the analysis thus far, which focused on the relationship between the propositions at issue and

higher-order propositions, including but not limited to the disposition of the case, from the analysis that follows, which focuses on whether the relationships between or among individual propositions should alter this initial presumption.

i. The problem of structured analysis

One argument for crediting nonsupportive propositions as holdings immediately arises, and this argument also will reveal a potential reason to discount some supportive propositions as dicta. Even though statement (6) is a nonsupportive proposition, the question of whether an interest is compelling is part of a legal test. We consider legal analysis to be *structured* when a legal test or legal doctrine, whether adopted in the case itself or in prior cases, identifies what a party must show to prevail on an issue. Strict scrutiny provides that to defend a certain type of program (in *Bakke*, programs that use racial classifications), the government must show both a compelling interest and that the program is narrowly tailored to that interest. When structured legal analysis identifies two or more elements, factors, or prongs as required to resolve a particular issue, there is an argument that a court's consideration of any one of these should count as a holding even if that portion of the analysis is nonsupportive, because consideration of elements of a structured analysis might not seem like a mere aside.⁷¹

With unstructured analysis, in contrast, the courts may consider various issues, but there is no legal test or doctrine that explicitly ties those issues together or requires their collective resolution. One might imagine a hypothetical *Bakke* decided in a world in which there was no such thing as strict scrutiny. The Court might still have been willing to entertain arguments about issues such as the importance of diversity and whether the Davis program was genuinely geared to advancing this interest. But those two considerations would not have resulted from any test or doctrine that identified them as the exclusive issues of interest or as related factors driving the result on one issue. While common law courts often apply doctrinal tests,⁷² they also often engage in case-by-case reasoning in which the import of any portion of the analysis is not specified in advance. Another example of unstructured analysis might be a case in which a litigant has made two distinct challenges to a statute, for example under the First Amendment of

⁷¹ We recognize that the phrase “a mere aside” is question-begging, and we will defend our conclusions about structured analysis more rigorously later. See *infra* Part III.C.4. Our immediate purpose is simply to identify a potential reason for deviating from the presumptive definition of holding.

⁷² For a defense of the use of doctrinal tests in constitutional adjudication, see Richard H. Fallon, Jr., *Foreword: Implementing the Constitution*, 111 HARV. L. REV. 54 (1997).

the U.S. Constitution and under a parallel state constitutional provision. In such a case, in contrast to the hypothetical *Bakke* without a formulated strict scrutiny test, the relationship of each issue to the disposition of the case would be clear. As with the *Bakke* hypothetical, however, there would be no test or doctrine tying the two issues together (after all, if the case had arisen in a state without such a constitutional provision, the state issue would not even have arisen), and so we would label the relationship between the two issues to be unstructured.⁷³

There is a further important distinction within the category of structured analysis, between ordered and unordered analysis. Even though Justice Powell's statement of the strict scrutiny test in statement (3) mentions the tailoring issue before the compelling interest issue, it might appear that the most natural order for strict scrutiny analysis is to consider the compelling interest issue first. Lawyers might intuit that one should resolve pure questions of law, such as whether an interest is compelling, before resolving applications of law to fact, such as assessing whether a particular program is narrowly tailored to an interest. Moreover, the narrow tailoring prong requires for its application at a minimum the assumption of an interest to which the program will be linked. In contrast, the compelling interest prong does require any understanding about the program that must be linked to that interest based upon narrow tailoring. Lawyers might therefore intuit, rightly or wrongly, that because the reverse of the conventional application of the test would require the following a three step analysis—(1) a presumed interest, (2) an assessment linking the program to the interest based upon narrow tailoring, and (3) a determination whether the assumed interest is compelling—the more logical approach is instead a straightforward two-step analysis in which we first assess whether the interest is compelling and only then assess narrow tailoring.

The judgment that the strict scrutiny test has a natural order, however, may be based as much on aesthetics as on logic. The following wording of a strict scrutiny test seems natural enough and would seem to imply a different order: "The state must prove its challenged law is narrowly tailored to serve a proffered governmental interest, and further establish that the

⁷³ As with most legal distinctions, there may be close cases between structured and unstructured analysis. Consider, for example, a plaintiff who challenges a statute on the basis that it violates the Fourteenth Amendment of the U.S. Constitution, with specific challenges based upon the Privileges or Immunities Clause, the Equal Protection Clause, and the Due Process Clause. This case is structured in the sense that one might say that constitutional doctrine clearly indicates that to find a violation of the Fourteenth Amendment, a court must find a violation of one of its clauses. Our tendency, however, is to see this as an example of unstructured analysis, because there is no doctrinal interdependence between or among the individual constitutional claims and because these issues are frequently, indeed most commonly, raised in isolation of one another.

proffered interest is compelling.” A requirement that a classification be tailored to serve a proffered governmental interest, which is then established as compelling seems to differ only in emphasis from Justice Powell’s requirement that a classification be narrowly “tailored to serve a compelling governmental interest.”⁷⁴ If forced to choose which of the strict scrutiny prongs comes first, most constitutional lawyers, we suspect, would choose the compelling interest prong. This intuition, however, appears to be largely a function of custom, and the fact that in the English language, the wording Justice Powell used happens to be more elegant and compact than the alternative. We thus recognize that lawyers may disagree about the most natural order of many legal tests, not limited to strict scrutiny, and about whether some legal tests are ordered at all.

Different tests and doctrines lie along a continuum with respect to how strongly or weakly ordered they are, although individual tests might prove difficult to classify. At one end of the continuum lies rigidly ordered doctrine, for which a court is absolutely required to consider one issue before another. The paradigmatic example is that courts must resolve certain threshold issues, such as standing, favorably to the claimant before considering the merits of a case.⁷⁵ If that is the right end of the continuum, then slightly to the left would be cases in which the authoritative decisionmaker has definitively indicated that there is a typical, or seemingly natural, order, but that in an extraordinary case, a court might approach issues in a different order. An example of this is the general principle that issues of subject matter jurisdiction should be resolved before issues of *in personam* jurisdiction, given that a defect in personal jurisdiction, unlike a defect in subject matter jurisdiction, can be waived. The Supreme Court has recognized this principle but decided that it is not absolute.⁷⁶ Loosely ordered tests like strict scrutiny are further to the left of the continuum. And to the far left are multifactor tests where there is no priority, or interdependence, among the various identified factors.⁷⁷

Regardless of whether analysis is unstructured, structured but unordered, or structured and ordered, in the course of its written opinion, a court will necessarily consider the issues that

⁷⁴ 438 U.S. at 299 (opinion of Powell, J.).

⁷⁵ See *Steel Co. v. Citizens for a Better Environment*, 523 U.S. 83 (1998) (rejecting the practice of hypothetical jurisdiction), *discussed infra* text accompanying notes x-x.

⁷⁶ See *Ruhrgas AG v. Marathon Oil Co.*, 526 U.S. 574 (1999) (finding that questions of subject matter jurisdiction ordinarily should, but need not always, analytically precede questions of personal jurisdiction), *discussed infra* notes x-x.

⁷⁷ For a discussion of one such test, see *infra* Part V.C.

it addresses in a chosen order. The ordering of issues within a judicial opinion presents an additional distinction potentially relevant to assessing whether nonsupportive propositions should be credited as holdings. Sometimes, by the time the court reaches a particular issue in its opinion, it will be clear, even placing aside any clues from an introductory roadmap that the case may provide, that because of the manner in which the court has resolved the other case issues, the resolution of the remaining issue cannot affect the case judgment regardless of how it is resolved. Although the issue may have been potentially supportive at various points in the case or on appeal, thus motivating counsel to advance arguments concerning its preferred resolution, it has become nonsupportive by virtue of the court's prior resolution of issues that proved controlling earlier in the opinion. Of course this is not always the case, and quite frequently, the court will resolve issues early on in an opinion in a manner that requires the resolution of other supportive issues to determine the ultimate judgment.⁷⁸ The distinction that we are drawing is whether an issue is or is not of continuing relevance, in the sense of potentially affecting the outcome, by the time a court addresses that issue in its opinion based upon the manner in which it has already resolved the preceding issues in the same opinion. When issues have continuing relevance at the point at which the court addresses them in its opinion, there is an argument for considering the resolution of those issues to constitute holdings, even though in the final analysis the resolution might prove nonsupportive.

Overlaying the distinction between continuing relevance and no continuing relevance onto the three-way distinction among unstructured, structured but unordered, and structured and ordered analyses produces six possible classifications for nonsupportive propositions. In addition to these six categories, it is possible for a proposition to be absolutely irrelevant in the sense that nothing would turn on its resolution regardless of the order in which the court considered it relative to the remaining issues discussed in the opinion. Table 1 places the seven resulting categories of nonsupportive propositions along a continuum representing the strength of the presumption favoring, or disfavoring, holding status. Absolutely irrelevant propositions present the strongest case for disfavoring holding status, or for presuming dicta. The case of an ordered test where the nonsupportive proposition remains relevant at the point of the analysis where the court considered it presents the strongest case for crediting nonsupportive propositions as

⁷⁸ A simple example: Finding that jurisdiction or standing is met invites resolution of the supportive issues addressing the merits of the case.

holdings. If a court must find *A* and then *B* in that order to rule for a party, and the court finds *A* but then concludes not *B*, there is a strong argument that *A*, even though it is ultimately a nonsupportive proposition, should count as a holding. Because there is a continuum representing the degree to which tests are or are not ordered, there is also a continuum from “structured but unordered, continuing relevance” to “structured and ordered, continuing relevance.” For reasons that we will elaborate later,⁷⁹ the closer a proposition on this continuum is to being a structured and ordered proposition with continuing relevance, the stronger will be its claim to the status of holding.

The examples in the table are based upon three hypotheticals. The first, the paradigm of ordered analysis, is a case in which because standing is jurisdictional, the court must resolve standing in favor of the claimant before it addresses the merits. The second, an example of an unstructured analysis, is a case in which a plaintiff challenges a statute banning the advertising of handguns on both First Amendment and Second Amendment grounds; to prevail the plaintiff needs to succeed on either ground. The third, an example of a structured but unordered analysis, is a case in which a plaintiff challenges a state’s refusal to honor a gay marriage from another state under the Full Faith and Credit Clause; to prevail, the plaintiff must establish that the marriage should count as a “public act” or as a “judicial proceeding” of the other state.⁸⁰ In all these examples, we assume that there are no other contested issues in the case.

⁷⁹ One way to think of this is to recognize that when a test is implicated in a proper case, random factors beyond the control of the court present the need to apply the test and, as a result, there is little risk that applying the entire test is the product of judicial manipulation. *See infra* Part III.C.4.

⁸⁰ *See* U.S. CONST. art. IV, § 1 (“Full Faith and Credit shall be given in each State to the public Acts, Records, and judicial Proceedings of every other State. And the Congress may by general Laws prescribe the Manner in which such Acts, Records and Proceedings shall be proved, and the Effect thereof.”). *See generally* Paige E. Chabora, *Congress’ Power Under the Full Faith and Credit Clause and the Defense of Marriage Act of 1996*, 76 NEB. L. REV. 604 (1997) (discussing the applicability of the Full Faith and Credit Clause to the question of gay marriage).

Table 1. Types of nonsupportive propositions⁸¹

	Category	Example
← Stronger claim to holding ←	(1) Absolutely irrelevant	In any of the three hypotheticals, the court mentions that Michael Jordan is the greatest basketball player ever. This could not affect the outcome of the case regardless of how the court resolved the other propositions, and would thus clearly have no precedential value in a subsequent case in which Jordan sues on a contract in which, for good consideration, he is promised a large bonus if he establishes a reputation as the greatest basketball player ever.
	(2) Structured and ordered, with no continuing relevance	In the threshold issue hypothetical, the court finds that the plaintiff does not have standing, but then rules that the plaintiff's argument on the merits is correct. The resolution of the merits issue is an ordered nonsupportive proposition that has no continuing relevance at the time the court considers the issue.
	(3) Unstructured, with no continuing relevance	In the handgun advertising hypothetical, the court strikes down the statute on First Amendment grounds, but then finds that the Second Amendment would not provide a basis for striking down the statute. The latter conclusion is a nonsupportive proposition that is part of an unstructured analysis and that has no continuing relevance by the time the court considers the issue.
	(4) Structured but unordered, with no continuing relevance	In the gay marriage hypothetical, the court finds that the marriage is a "public act," and is thus entitled to Full Faith and Credit, and then concludes that the marriage would not count as a "judicial proceeding." The latter conclusion is a nonsupportive proposition that is part of a structured, but unordered analysis and that has no continuing relevance by the time the court considers the issue.
	(5) Unstructured, with continuing relevance	In the handgun advertising hypothetical, the court rejects the First Amendment argument, but then strikes down the statute on Second Amendment grounds. The former conclusion is a nonsupportive proposition that is part of an unstructured analysis, but that had continuing relevance at the time the court considered the issue.
	(6) Structured but unordered, with continuing relevance	In the gay marriage hypothetical, the court finds that the marriage is not a "public act," but then concludes that the marriage was a "judicial proceeding," and is thus entitled to Full Faith and Credit Clause protection. The conclusion with respect to "public act" is a nonsupportive proposition that is part of a structured but unordered analysis, but that had continuing relevance at the time the court considered the issue.
	(7) Structured and ordered, with continuing relevance	In the threshold issue hypothetical, the court finds that the plaintiff has standing but then rules against the plaintiff on the merits. The former conclusion is a nonsupportive proposition that is part of an ordered analysis that had continuing relevance at the time the court considered the issue.

⁸¹ This table arranges nonsupportive propositions along a continuum, from weakest presumptive claims to holding status at the top to strongest presumptive claims to holding status at the bottom. The broken line between rows (4) and (5) signifies that while we believe that there is a stronger claim to holding for propositions with no continuing relevance but that form part of a structured and unordered analysis, we recognize that others might instead view the distinction between continuing relevance and no continuing relevance as more important than that between structured and unstructured, in which case a proposition that is part of an unstructured analysis with continuing relevance would be entitled to a relatively stronger presumption of holding status.

So far, we have considered the problem of ordered analysis only with respect to nonsupportive propositions. The problem of ordered analysis, however, may also imply that some supportive propositions, which presumptively count as holding, should instead be discounted as dicta. The analysis is simpler for supportive propositions, however, because there is only one category—structured and ordered, with no continuing relevance—that presents an argument for converting holding to dicta.⁸² Where a test is ordered, and the court resolves the first prong of the test in a way that determines the outcome under the test as a whole, there is a strong argument that the court’s resolution of the second prong should be discounted as dicta even if it would otherwise count as an alternative justification and a holding. Suppose, for example, that a court concluded that it did not have jurisdiction and yet proceeded to resolve the merits.⁸³ If the court resolved the merits against the petitioner,⁸⁴ the resolution of the merits might be seen as an alternative justification for the petitioner to lose. Nonetheless, the problem of structured and ordered analysis would provide an argument for discounting the merits resolution as dicta even if alternative justifications are generally holdings.

ii. *The problem of hypotheticals*

The analysis so far has provided an interpretation of Powell’s argument that would make statement (8) (a plus factor plan would not hinder the attainment of diversity) a supportive proposition, and thus applying our definition, presumptively a holding.⁸⁵ Even so, there may be another argument for denying holding status to statement (8). The court’s analysis of the Harvard plan is counterfactual. Statement (8) can be interpreted as a conditional statement: “If a plan has the characteristics of the Harvard plan, then it will not hinder achievement of genuine diversity.” The antecedent of that statement—“a plan has the characteristics of the Harvard plan”—is false

⁸² The other categories at most simply restate the problem of alternative justifications. Consider, for example, supportive propositions in unstructured analysis with no continuing relevance. In a variation on the handgun advertising hypothetical depicted in row (3) in Table 1, suppose the court strikes down the statute on First Amendment grounds, but now finds that the Second Amendment provides an additional basis for striking down the statute. The Second Amendment conclusion is a supportive proposition that is part of an unstructured analysis, but that has no continuing relevance at the time the court considers the issue. Whether this proposition should count as a holding or dicta depends entirely on the problem of alternative justifications.

⁸³ See *infra* text accompanying notes x-x (discussing hypothetical jurisdiction).

⁸⁴ For an infamous illustration of a case fitting this paradigm, see *Dred Scott v. Sandford*, 60 U.S. (19 How.) 393 (1857), where having found no diversity jurisdiction, because Dred Scott was not a citizen of a state for purposes of diversity jurisdiction under Article III of the United States Constitution, the Court nevertheless proceeded to reach the merits of the case, striking down the Missouri Compromise. For a critical analysis of the first holding, see Cass R. Sunstein, *The Dred Scott Case*, 1 GREEN BAG 2D 39 (1997).

⁸⁵ See *supra* notes 64-70 and accompanying text.

in *Bakke*, because the Davis plan did not have the characteristics of the Harvard plan. If a Harvard-like plan were presented in a subsequent case, a lawyer challenging the program might argue that even if statement (8) formed a part of Powell's *Bakke* analysis in a manner that ordinarily would entitle it to presumptive holding status, the counterfactual nature of the statement should suffice to rebut that presumption.

We will label this the problem of *hypotheticals*. Reasoning about actual cases based upon hypothesized case facts implicating significantly distinguishable cases yet to be presented is not only not unusual, it is the standard fare of judging. Judges frequently pose hypotheticals about undecided cases in oral arguments, and analogical reasoning is frequently reflected in published opinions.⁸⁶ A judge in Case *A* might conclude that Case *A* is sufficiently similar to Case *B*, a case with different facts, that they should produce the same result. That poses no problem for our purposes if the outcome in Case *B* is already clear by virtue of a prior holding. It might pose a problem, however, if the outcome in Case *B* is not resolved, and the judge in Case *A* anticipates a result in Case *B* to facilitate its resolution in the case before it.⁸⁷ Does an anticipated result of Case *B* count as a holding about how to resolve those facts, if the anticipation itself forms a sufficiently integral part of the reasoning that, were it not counterfactual, it would count as a holding of Case *A*? Analogical reasoning is undoubtedly important in allowing judges to sharpen their analyses of factual and legal issues presented on appeal, but this does not necessarily mean that analogical reasoning creates additional holdings.⁸⁸

⁸⁶ Assessments of the usefulness of analogical reasoning in judicial opinions include Scott Brewer, *Exemplary Reasoning: Semantics, Pragmatics, and the Rational Force of Legal Argument by Analogy*, 109 HARV. L. REV. 923 (1996); Emily Sherwin, *A Defense of Analogical Reasoning*, 66 U. CHI. L. REV. 1179 (1999); and Cass R. Sunstein, *On Analogical Reasoning*, 106 HARV. L. REV. 741 (1993).

⁸⁷ As an example, suppose a court faces a case involving whether flag burning should receive First Amendment protection. The court might reason that flag burning should be treated like cross burning, and even though the court might never have considered cross burning, the court might conclude that cross burning should receive First Amendment protection and therefore that flag burning therefore should as well. (For consideration of the flag burning-cross burning analogy in the analogical reasoning literature, see Sunstein, *supra* note 86, at 759-61.) The question is whether this creates a holding as to cross burning, should that issue subsequently come before the court.

⁸⁸ The analogical reasoning in statements (7) and (8) arguably differ from conventional analogical reasoning, because the statements are being used to address legal questions the resolutions of which arguably *require* a form of analogical reasoning. An analysis of the narrow tailoring prong of strict scrutiny demands the identification or rejection of a more narrowly tailored alternative. See *supra* note 65. Thus, use of analogical reasoning may be necessary rather than merely sufficient to establish some higher-order proposition. In the end, though, this argument is a nonstarter. Even if the court must use analogical reasoning to assess a particular proposition, there would be no requirement that it use any particular analogy.

iii. *The problem of biconditional statements*

The analogical reasoning problem presents a possible reason for counting statement (8) (a plus factor plan would not hinder the attainment of diversity) as dicta even if it is a presumptive holding, and statement (8) also must count as presumptive dicta if interpreted as a nonsupportive proposition.⁸⁹ Either way, the problem of biconditional statements provides an argument for counting statement (8) as holding. The argument is that even if the propositions related to the Harvard plan were irrelevant in the formal sense that they did not lead to the case result, they were relevant in the sense that they articulated a resolution of the central issue of the case. That issue is the extent to which it is acceptable for government educational institutions to use race as a consideration. In this sense, endorsement of statement (8) and other statements concerning the Harvard plan were indeed supportive of the court's resolution of this issue and were not hypothetical.

Justice Powell himself came close to justifying the Harvard discussion on this basis. Countering a complaint⁹⁰ from Justice Stevens that his opinion had gone further than necessary to resolve the case,⁹¹ Powell noted that the lower court had rejected the Davis plan on the broad ground that Davis had considered race at all, rather than on the potentially narrower ground that Davis had improperly employed race in the form of a quota.⁹² The reference to the lower court might obscure the point. Whether a statement in a Supreme Court opinion is holding or dicta depends on the issue presented to the Supreme Court, not on the issues that the lower court

⁸⁹ See *supra* text accompanying note 67.

⁹⁰ Justice Stevens's criticism that Justice Powell engaged in an unnecessary discussion demonstrates that the distinction between holding and dicta not only affects what future courts will consider authoritative in a precedent-setting opinion, but also might affect the precedent-setting court's own determination of how much to say. Thus, while some jurists might willingly defend against accusations of claiming holding status for what is arguably dicta, as Powell appears to have done regarding statement (5), or happily admit to authoring dicta without making this stronger claim, others might see dicta as something that is best avoided. See, e.g., *Hollister Convalescent Hosp., Inc v. Rico*, 542 P.2d 1349, 1356-58 (Cal. 1975) (referring to "unnecessary and overbroad dicta," "ill-considered dicta," "erroneous dicta," "panoramic dicta," and "persistent dicta"); see also *infra* Part III.B (arguing that opinions closely tied to the material facts of a case may be normatively preferable to opinions that stray from the material facts).

⁹¹ Justice Stevens noted that, following the California Supreme Court ruling, there was no "outstanding injunction forbidding any consideration of racial criteria in processing applications." 438 U.S. at 410-11 (Stevens, J., concurring in the judgment in part and dissenting in part). Justice Stevens concluded: "It is therefore perfectly clear that the question whether race can ever be used as a factor in an admissions decision is not an issue in this case, and that discussion of that issue is inappropriate." *Id.* at 411. For a critique of Powell's opinion along these lines, see Arval A. Morris, *The Bakke Decision: One Holding or Two*, 58 OR. L. REV. 311, 326-32 (1979).

⁹² 438 U.S. at 271 n.** (opinion of Powell, J.) (stating that the California Supreme Court opinion "left no doubt that the reason for its holding was petitioner's use of race in consideration of *any candidate's* application"). Justice Powell also emphasized that the medical school "cross-complained in the trial court for a declaratory judgment that its special program was constitutional." *Id.* That a party sought declaratory judgment in a suit, however, ordinarily would not be thought of as allowing the courts to consider issues beyond the controversy between those parties. See, e.g., *Md. Cas. Co. v. Pac. Coal & Oil Co.*, 312 U.S. 270, 273 (1941) (noting that jurisdiction in declaratory judgment suits is limited to the actual controversy between the parties).

happened to discuss.⁹³ If, for example, the lower court had also explored an argument by Bakke that he was the victim of sex discrimination, but the Supreme Court failed to grant certiorari on this basis, that issue would not be before the Court. Or, if the lower court had itself reached propositions that count as dicta, the lower court's decision to do so would not make the Supreme Court's decision to repeat or refute those statements holdings. The Supreme Court itself has stated that it "reviews judgments, not statements in opinions."⁹⁴

Conceived in its best light, Powell's point is not so much that the issue of whether race could be used was resolved in the lower court, but that the issue of whether race could be used was before the Supreme Court itself.⁹⁵ Powell's analysis of the Harvard plan constituted part of his resolution of this question. Let us now imagine that Powell had organized his opinion differently. Powell might have started with the question whether race constitutionally could be considered in admissions decisions, concluding that such consideration was acceptable as long as race was used as a plus factor rather than as a quota, and then considered how his resolution applied to the facts of the Davis plan, concluding that the Davis plan counted as an illicit quota. Even if the Harvard plan were used as an example to explain the plus factor concept, the conclusion that race can be used as a consideration and the distinction between a plus factor and a quota no longer would seem to be as much of a detour.

Our point is not to criticize the internal organization of Powell's opinion, but rather to recognize that in that opinion, Justice Powell effectively created a test for whether use of race in admissions decisions is constitutional. While subsequent commentators unmistakably extrapolated the test from Powell's analysis,⁹⁶ surprisingly, the observation that it was a test did not appear to factor into anyone's analysis of whether the discussion of the Harvard plan was

⁹³ See *Morris*, *supra* note 91, at 331 ("If each reason given by a lower court in support of its judgment constituted an issue properly before the United States Supreme Court, the consequences would be chaos."). A higher court can choose not to pass on many propositions made by a lower court. A reviewing court, reaching the same result on a narrower ground as the court below, has elected to hold off resolution on the broader ground unless and until a case arises that disallows resolution on the selected narrower ground and that instead demands resolution—in either direction—on the broader ground. Justice Powell's implication that he was forced to reach the issue because the lower court mentioned it thus seems unpersuasive. Whether or not he was compelled to do so, however, does not answer whether Powell overreached in electing to reach the issue arising at the broader level, specifically whether the use was altogether impermissible.

⁹⁴ *Black v. Cutter Labs.*, 351 U.S. 292, 297 (1956).

⁹⁵ These are admittedly connected. The issue was before the Supreme Court in part because it had previously been resolved in the lower court, and also because the Court had granted certiorari on that issue.

⁹⁶ See, e.g., Paul J. Mishkin, *The Uses of Ambivalence: Reflections on the Supreme Court and the Constitutionality of Affirmative Action*, 131 U. PA. L. REV. 907, 924 (1983) (noting that Justice Powell distinguished between programs that used race as a plus factor and those that used race to justify quotas).

holding or dicta.⁹⁷ Tests, of course, are ubiquitous in law,⁹⁸ and we have already seen that statement (3) (all race classifications are subject to strict scrutiny) adopted a test to be applied to affirmative action plans. A court's decision to formulate a test may present two distinct problems, which we will refer to as the issue of *biconditional statements* and the issue of *breadth*.

We will first consider biconditional statements. If we accept the proposition that equal protection condones some benign use of race, we might imagine a boundary that divides permissible from prohibited methods of using race. According to the *Bakke* Court, the Davis quota plan was on the wrong side of that line. Powell could have made clear that the Davis program was on the wrong side of the line without making clear exactly where the line was.⁹⁹ For example, he might have written, "A quota will never be narrowly tailored to the goal of improving diversity. Whether a race classification other than a quota, such as Harvard's plus factor plan, can be narrowly tailored is a question for another day."

Such an opinion would have endorsed the proposition "if quota, then not narrowly tailored," but it would have said nothing about the inverse¹⁰⁰ proposition "if not quota, then narrowly tailored" (or, what amounts to the same thing, "if instead a plus factor approach, then narrowly tailored"). Justice Powell's opinion, however, endorsed both the proposition, in statement (7) (focusing exclusively on race would hinder genuine diversity), and its inverse, in statement (8) (a plus factor plan would not hinder attainment of diversity). The test that Powell created is thus the biconditional statement "if and only if a quota, then not narrowly tailored." The biconditional statement issue is whether this statement counts as a holding, assuming that statement (7) would count as a holding, in which case statement (8) should count as a holding as well.

⁹⁷ See *infra* Part II.A.3 (discussing the prior literature assessing the holding-dicta distinction in *Bakke*).

⁹⁸ See Fallon, *supra* note 72, at 67-73 (offering a typology of eight different kinds of tests in constitutional law).

⁹⁹ Those who joined the Stevens opinion in *Bakke* would have held that *any* use of race violated Title VI of the Civil Rights Act of 1964, and thus indicated that the Davis plan was far from the boundary line for the permissive use of race. See *supra* note 32. Justice Powell, in contrast, elected to inform courts and relevant state actors that in his view, the Davis plan was sufficiently close to the boundary line that with some modifications, Davis could salvage its affirmative action program. Had Powell been unable to make that proffer, it is at least conceivable that he might have preferred to endorse the Davis plan in full, rather than risk signaling that any use of race—including the Harvard plan which he endorsed—was illicit. See *infra* note x and accompanying text.

¹⁰⁰ The implication "if not *p*, then not *q*" is the inverse of the statement "if *p*, then *q*." The inverse is thus the contrapositive of and logically equivalent to the converse (if *q*, then *p*). See, e.g., <http://www.jimloy.com/logic/converse.htm> (last visited September 2, 2004) (discussing the definition of converse, inverse, and contrapositive).

iv. *The problem of breadth*

Powell's endorsement of the quota test also presents the related but analytically separate problem of breadth. Powell might have described the facts of the Davis plan, and concluded, based upon the aggregation of those facts, that the program failed narrow tailoring. This presumably would have produced a holding—"if a program is just like the Davis program, then it is not narrowly tailored"—but the holding would be extremely limited. That might be the narrowest possible proposition that would support the judgment, but Powell also might have made some other generalizations in between this level of generality and the level of generality that he ultimately chose. For example, Justice Powell might have singled out schools that have racial quotas but that do not also ensure a minimum number of places for applicants with other qualities.¹⁰¹ Alternatively, Powell might have endorsed only the proposition that "if a race-based quota is used in a graduate admissions program, then it is unconstitutional," leaving aside the question of race-based quotas at other educational institutions. In sum, Powell could have chosen narrower reasoning than he apparently selected without affecting the ultimate case judgment. The question is whether a judge's decision to endorse a broader proposition than necessary in resolving a case makes the broader selected proposition dicta. We will eventually conclude that it does not.¹⁰²

The breadth issue affects whether a later court may distinguish a precedent from an earlier court. A later court cannot distinguish a prior case in a way that is contrary to any of that case's holdings. If Justice Powell had issued only the narrowest possible ruling, focusing only on the facts, then a later court could have distinguished *Bakke* easily. For example, a court might have concluded that a program just like the Davis plan, but with a smaller percentage of spaces reserved for minority students, was acceptable. Given *Bakke* as written, a later court distinguishing *Bakke* on the same ground would need to conclude, explicitly or implicitly, either

¹⁰¹ Powell disliked the Davis program because racial minorities were singled out for special treatment, unlike those with "exceptional personal talents, unique work or service experience, leadership potential, maturity, demonstrated compassion, a history of overcoming disadvantage, ability to communicate with the poor, or other qualifications deemed important." 438 U.S. at 317 (opinion of Powell, J.). Powell's opinion suggests that it might have been sufficient to sustain the program had Davis singled out additional groups for reserved places that are not open to the general applicant pool. Whether or not this is a satisfactory reading of *Bakke*, the analysis reveals that an author of an opinion must select from a continuum of breadth.

¹⁰² As we will later demonstrate, *see infra* notes x-x and accompanying text, an understanding of holding that would relegate to the status of dicta any rationale that is broader than necessary in resolving a particular case operates in tension with the Supreme Court's narrowest grounds rule articulated in *Marks v. United States*, 430 U.S. 188, 192 (1973). Such an understanding would suggest that the Court's holding is limited to the narrowest possible grounds even when a majority, or indeed a unanimous Court, stands united behind a broader rationale in support of the ultimate case disposition.

that Powell’s reasoning did not endorse the proposition that racial quotas are unconstitutional or that the proposition was dicta. Otherwise, the distinguishing court would seem disingenuous. Our analysis cannot do all the work of clarifying when it is permissible for a later court to distinguish a prior case, of course. That will also depend on resolution of ambiguities about what propositions the precedent-setting court indeed endorsed.¹⁰³

If unnecessary breadth makes a statement dicta, then both statements (7) (focusing exclusively on race would hinder genuine diversity) and (8) (a plus factor plan would not hinder the attainment of diversity) would be dicta because the biconditional statement analysis above suggested that statement (8) might be a holding only on the assumption that statement (7) counts as a holding. Let us, however, assume (as we will later conclude¹⁰⁴) that broad statements do not make statements dicta. The final statement other than the disposition, statement (9) (the Harvard plan would be constitutional), is easily analyzed. Statement (9) followed logically only because Powell had already endorsed statement (6) (diversity is a compelling interest).¹⁰⁵ Against the backdrop of the remaining statements, given either statements (6) or (9), the other statement follows as a corollary. Where a statement’s truth depends on that of other statements, it logically can count as a holding only if the other statements count as holdings.

3. Summary and Comparison with Previous Analyses

We have now completed the necessary analytical apparatus with which to analyze Justice Powell’s *Bakke* opinion. Table 2 includes in the first column each of the propositions discussed above. The second column states whether the identified proposition is necessary or sufficient to some higher-order proposition in the case, and the third column states whether this classification makes it a supportive proposition, thus warranting presumptive holding status, or a nonsupportive proposition, thus warranting presumptive dicta status. Finally, the fourth column identifies the reasons that the presumptive status accorded each proposition in the third column might be rebutted, thus discounting some presumptive holdings as dicta and crediting some

¹⁰⁴ See *infra* Part III.C.7.

¹⁰⁵ It might appear that statement (9) also presents the problem of breadth, because the court concludes that a plus factor approach would be constitutional, rather than merely drawing the narrower conclusion that a plus factor approach is narrowly tailored. But this issue is analytically distinct from the problem of breadth. Consider statement (9) in conditional form: “if a program uses a plus factor approach, then constitutional.” The problem of breadth concerns the antecedent to this conditional (a program uses a plus factor approach), not the consequent. Whether the breadth of the consequent should count as a holding depends on whether statement (6) counts as a holding.

Deleted: ¶

presumptive dicta to holding. Because statement (8) is ambiguous,¹⁰⁶ we have offered two constructions, each with its own analysis.

Table 2. The *Bakke* Propositions and their Presumptive and Ultimate Holding-Dicta Classifications

Proposition	Classification of proposition	Presumptive status	Possible rationales for rebutting presumptive holding or dicta status
(1) Davis program is race-based.	Necessary to statement (10)	Holding	None
(2) Equal protection protects nonminorities.	Necessary to statement (10)	Holding	None
(3) All race classifications are subject to strict scrutiny.	Test; selection of some test was necessary to statement (10)	Holding	Problem of alternative possible justifications; the Court might have instead concluded that the Davis program failed a laxer test.
(4) Remedying the present effects of past societal discrimination is not a compelling interest.	Sufficient for proposition that goal of remedying the present effects of past societal discrimination does not pass strict scrutiny	Holding	Problem of alternative possible justifications; the same result could be achieved under strict scrutiny by refuting narrow tailoring.
(5) Improving health services for underprivileged communities was not shown to be a goal to which program was narrowly tailored.	Sufficient for proposition that goal of improving health services for underprivileged communities does not pass strict scrutiny	Holding	Problem of alternative possible justifications; the same result could be achieved under strict scrutiny by finding no compelling state interest.
(6) Diversity is a compelling interest.	Nonsupportive proposition	Dicta	Problem of structured analysis; the compelling interest prong is a part of the structured strict scrutiny analysis and arguably is ordered first in that analysis.
(7) Focusing exclusively on race by means of a quota would hinder genuine diversity.	Sufficient for proposition that Davis program is not narrowly tailored to goal of achieving diversity	Holding	(1) Problem of alternative justifications; the second interpretation of statement (8) is also sufficient for the proposition that the Davis program is not narrowly tailored to goal of achieving diversity. (2) Problem of breadth; Powell could have issued a narrower holding.
(8) A plus factor plan would not hinder attainment of diversity.	Arguably nonsupportive proposition	Dicta	Problem of biconditional statements, assuming that (7) is a holding; Powell established the quota vs. plus factor dichotomy as the means of determining narrow tailoring.

¹⁰⁶ See *supra* text accompanying notes 65-66.

Proposition	Classification of proposition	Presumptive status	Possible rationales for rebutting presumptive holding or dicta status
	Arguably sufficient for proposition that Davis program is not narrowly tailored to goal of achieving diversity	Holding	(1) Problem of alternative justifications; statement (7) is also sufficient for the proposition that the Davis plan is not narrowly tailored to goal of achieving diversity. (2) Problem of breadth; Powell did not have to resolve the constitutional status of all plus factor plans. (3) Problem of hypotheticals; the plan at issue was not a plus factor plan. (Note: If any of these three is adequate to convert the presumptive status, then the problem of biconditional statements might convert the status to dicta.)
(9) The Harvard plan would be constitutional.	Nonsupportive proposition	Dicta	As corollary of (6) and (8)
(10) The Davis plan is unconstitutional.	The disposition	Holding	None

Determining the status of Justice Powell’s discussion concerning the Harvard plan, and specifically whether any conclusions reached were dicta or holding, had significant implications for lower court litigation about affirmative action. And yet, those courts never adequately confronted the governing definitional issues. The Eleventh Circuit, for example, concluded, “We do not believe that Justice Powell’s opinion is binding, and his discussion of the Harvard Plan was entirely dicta.”¹⁰⁷ While that court rested the first conclusion on its analysis that the *Marks* narrowest grounds doctrine cannot be applied to *Bakke*,¹⁰⁸ the more important point for present purposes is that it offered no justification at all for the second conclusion.¹⁰⁹ The Sixth Circuit noted the Eleventh Circuit’s conclusion but declined to address whether the Eleventh Circuit was correct. “Even if this portion of Justice Powell’s opinion could be labeled dicta, it is nevertheless dicta from the determinative opinion in the only Supreme Court case to address the consideration of race and ethnicity in academic admissions,” the court observed, concluding that it therefore carries “considerable persuasive authority.”¹¹⁰

¹⁰⁷ Johnson v. Bd. of Regents of the Univ. of Ga., 263 F.3d 1234, 1261 (11th Cir. 2001).

¹⁰⁸ *Id.* at 1246-47. *But see supra* note 38 (arguing that the Powell opinion necessarily is controlling under *Marks*).

¹⁰⁹ The district court offered only a slightly more detailed justification of this position. “Powell’s view as to the validity of a ‘Harvard-style’ admissions system was mere dicta and not a holding in any event, since *Bakke* concerned a dual-track program rather than a ‘plus factor’ program like Harvard’s or UGA’s.” Johnson v. Bd. of Regents of Univ. System of Ga., 106 F. Supp. 2d 1362, 1369 (S.D. Ga. 2000), *aff’d*, 263 F.3d 1234 (11th Cir. 2001).

¹¹⁰ Grutter v. Bollinger, 288 F.3d 732, 746 n.9 (2002) (en banc), *aff’d*, 2003 WL 21433492 (U.S. June 23, 2003) (No. 02-241).

Legal scholars have considered the question whether the discussion of the Harvard plan should be treated as dicta in greater detail, but even those discussions have not identified the critical issues that we have shown here. In an informative discussion, Alan Meese argues, “The language approving the use of a plus system . . . was plainly gratuitous, dealing, as it did, with an issue—and an admissions program—that was not before the Court. This language was dicta under the conventional definition.”¹¹¹ Meese is correct that the Harvard plan was not directly before the Court, and perhaps that alone might be viewed by some as sufficient to remove Powell’s conclusions concerning plus-factor plans from the category of holding. But as demonstrated above, one can argue that Powell’s discussion of the Harvard plan was not gratuitous at all. Rather, the discussion was arguably essential to the narrow tailoring analysis of the Davis plan and within the confines of that discussion, merely constituting the “and only if” portion of his proffered biconditional holding on narrow tailoring. We will assess later whether that should count as enough, and while Meese has accurately described a “conventional definition” of dicta insofar as one exists, his lack of attention to this point of logical structure reflects the broader lack of attention in the scholarly literature.

The status of Justice Powell’s discussion was particularly critical to lower court litigation, because once lower courts identify a Supreme Court holding, they must follow it under the rule of vertical stare decisis.¹¹² The definitions of holding and dicta, however, also matter to the application of horizontal stare decisis, both at the circuit court level (where previous decisions are binding except in the unusual case of a prior panel decision that is vacated pursuant to en banc reconsideration or an intervening Supreme Court case that demonstrates that the prior panel case rests upon discredited precedent),¹¹³ and in the Supreme Court. In the recent Supreme Court

¹¹¹ Alan J. Meese, *Reinventing Bakke*, 1 GREEN BAG 381, 383 & n.12 (1998) (citing *Black’s Law Dictionary’s* definition of dicta). Meese’s application of the conventional definition seems reasonable, subject to the caveat that at least to Powell, the evaluation of the Harvard program may in fact have been necessary. See *supra* note 64 and accompanying text. We will argue, however, that the conventional definition is flawed and inconsistent with conventional practice. See *infra* Part IV.A.3.

¹¹² Under *State Oil Co. v. Khan*, 522 U.S. 3, 20 (1997) and *Rodriguez de Quijas v. Shearson/American Express, Inc.*, 490 U.S. 477, 484 (1989), lower federal courts and state courts are prohibited from anticipatorily overruling Supreme Court precedents based upon subsequent case law developments. Instead, lower federal courts and state court must wait until an express overruling by the Supreme Court itself. If one accepts the premise that both halves of Powell’s *Bakke* opinion—striking Davis’s use of race, while allowing some use of race in the form of a plus factor—expressed the holding under the narrowest grounds rule, then *Hopwood v. Texas*, 732 F.2d 932, 968 (5th Cir. 1996), which struck down the University of Texas Law School affirmative action program based largely upon post-*Bakke* developments in the Supreme Court’s race jurisprudence, violated the principle expressed in *Khan* and *de Quijas*. For further discussion of *Hopwood*, see *infra* notes 312-318.

¹¹³ See *supra* note x and accompanying text.

affirmative action litigation,¹¹⁴ most of the briefs ignored the issue.¹¹⁵ Perhaps the absence of discussion is attributable to a strong intuitive sense that Justice Powell's discussion of the Davis plan was dicta. The failure of affirmative action's opponents even to make any rigorous attack on the characterization of the Harvard plan as holding, however, reflects problems with existing definitions of dicta and the failure of commentators more broadly to consider the definitional question with precision.

Admittedly, the distinction should have mattered less than it would have had the relevant portion of Justice Powell's opinion been set out in a majority opinion. Had the Court wished to deviate from Powell's conclusions, it could have simply observed that whether or not the Powell opinion resolved *Bakke* on the narrowest grounds, the Supreme Court, unlike lower federal courts and state courts, need adhere only to precedents established in its own majority decisions.¹¹⁶ And while Justice O'Connor, writing for the majority in *Grutter*, adopted more or less in full Justice Powell's *Bakke* analysis, she did so while claiming no need to resolve whether Powell's *Bakke* opinion was a binding precedent under the narrowest grounds rule.¹¹⁷ After all, whether a statement is holding or dicta does not matter much if the subsequent court concurs with the merits of the stated proposition. Yet, at various points in her opinion, O'Connor justified her conclusions as much by citing Powell as by developing her own argument.¹¹⁸ Such citations might have seemed somewhat more persuasive if Powell's statements were clearly holdings (even if holdings not entitled to *stare decisis* weight), and conversely, somewhat less persuasive if Powell's statements were clearly dicta.

¹¹⁴ See *Grutter v. Bollinger*, 539 U.S. 306 (2003); *Gratz v. Bollinger*, 539 U.S. 244 (2003).

¹¹⁵ For a notable exception, see Brief Amicus Curiae of the Criminal Justice Legal Foundation in Support of Neither Party at 13-15, 2003 WL 164165, *Grutter* (No. 02-241). This brief concludes that Powell's discussion of the Harvard plan is dicta, based on a definition offered by Arthur Goodhart. *Id.* We will critique that definition later. See *infra* Part IV.A.2.

¹¹⁶ See *supra* note 39. This approach would have served the dual purposes of enabling the Supreme Court both to extricate itself from Powell's idiosyncratic approach to affirmative action and to remind lower federal courts and state courts of their continuing obligation to abide all Supreme Court precedents, including narrowest grounds decisions, unless and until the Supreme Court expressly overturns them, see *supra* note 112.

¹¹⁷ Justice O'Connor stated that it was unnecessary to perform the *Marks* inquiry, because "we endorse Justice Powell's view that student body diversity is a compelling state interest that can justify the use of race in university admissions." *Grutter*, 539 U.S. at 325. O'Connor gave no indication that she even recognized the existence of an issue of whether Justice Powell's comments on the Harvard plan were dicta.

¹¹⁸ See, e.g., *id.* at 337 ("The importance of this individualized consideration in the context of a race-conscious admissions program is paramount.") (citing *Bakke*, 438 U.S. at 318 n.52 (opinion of Powell, J.), without offering any additional defense).

B. Summary of Identified Problems

We have now created a presumptive definition of holding embracing each statement that is a supportive proposition, in the sense that it is necessary or sufficient either for the disposition or for another holding. We now will summarize the problem categories that, assuming that this is the appropriate presumptive definition,¹¹⁹ might reverse the presumptive classification. Table 3, which follows, identifies the problems that might turn presumptive holdings into dicta, or that might turn presumptive dicta into holdings, along with explanations and illustrations from the preceding discussion. Because the problem of structured analysis may have both effects, it is listed twice.

Table 3. Potentially presumption-altering categories, with explanations and illustrations

Potentially Presumption-Altering Category	Explanation	Example from <i>Bakke</i>
Alternative possible justifications	<i>A</i> and <i>B</i> are propositions that, if accepted, would each be sufficient for a holding. The court resolves and accepts <i>A</i> but does not consider <i>B</i> . The question is whether the failure to consider <i>B</i> denies <i>A</i> the status of holding.	Justice Powell found that the goal of remedying past societal discrimination was not a compelling interest, but he did not consider whether the Davis program was narrowly tailored to this goal. A negative answer to this question also would have been sufficient for the proposition that the goal of remedying past societal discrimination could not justify the Davis plan. The question is whether the finding that remedying past societal discrimination was not a compelling interest is a holding.
Alternative justifications	<i>A</i> and <i>B</i> are propositions that, if accepted, would each be sufficient for a holding. The court resolves and accepts both <i>A</i> and <i>B</i> . The question is whether <i>A</i> and <i>B</i> must both count as dicta because neither was necessary for resolution of the case.	Justice Powell arguably offered two justifications for the proposition that the Davis program failed narrowed tailoring: first, that the Davis program itself hindered genuine diversity; and second, that a more narrowly tailored program, the Harvard plan, would meet the constitutional requirement. The question is whether each finding negates the status of the other as holding.

¹¹⁹ See *infra* Part III.D.1-3 (arguing in support of this presumptive definition). If we determined that the problem of alternative possible justifications and the problem of alternative justifications were both sufficient to turn a holding into dicta, then we would need to abandon the presumptive definition.

Hypotheticals	A is a conditional proposition with an antecedent that is false on the facts of the case. The question is whether A should count as dicta as a result.	Justice Powell determined that if an affirmative action plan were structured like the Harvard plan, it would be narrowly tailored to the goal of improving diversity. (Note that even if the problem of hypotheticals would make this statement dicta, the problem of biconditional statements might restore this particular example to holding status.)
Breadth	A is a proposition that is a presumptive holding, but A is broader than necessary for the court to reach the disposition. The question is whether the broader than necessary portion of the statement is dicta.	Justice Powell concluded that if a program uses quotas, it is not narrowly tailored to the goal of improving diversity. The Court could have offered a narrower proposition, such as if a graduate educational program uses quotas or if it uses quotas based upon a particular percentage, it is not narrowly tailored to the goal of improving diversity.
Structured analysis (for presumptive alternative holdings)	A and B are propositions that are tied together by a doctrinal test. That test may be ordered or unordered. For example, suppose a legal test that the court is applying considers A before B. The court resolves A in a way that resolves the test as a whole (either finding A true if the test is "A or B" or finding A false if the test is "A and B"), and it then proceeds to resolve B in the same way. The question is whether the resolution of B should count as dicta, assuming that the problem of alternative justifications does not ordinarily turn presumptive holdings into dicta.	In a hypothetical alternative to <i>Bakke</i> , the Court concludes that Bakke did not have standing and that he would have lost anyway because affirmative action is constitutionally permissible. Even if the conclusion that affirmative action is constitutionally permissible ordinarily would count as a holding because alternative justifications in fact do count as holdings, the question is whether the proposition should count as dicta since the first prong of the analysis (standing) made it unnecessary for the Court to consider the second prong (merits).
Biconditional statements	A is a holding, and B, which is not a presumptive holding, is the inverse of A. The question is whether B's status as the inverse of A elevates B to the status of holding.	Justice Powell identified as the test for meeting narrow tailoring to the goal of diversity whether the program took the form of a plus-factor plan (permissible) or a quota (impermissible). Assume that the following counts as holding: If a program uses quotas, it is not narrowly tailored to the goal of improving diversity. Now assume that biconditional statements count as holdings. Then, the following would also count as a holding: A program taking the form of a plus factor plan is narrowly tailored to further the goal of academic diversity.
Structured analysis (for presumptive dicta)	A legal test that the court is applying considers A before B. The court resolves A in a way that does not resolve the test as a whole (either finding A false if the test is "A or B" or finding A true if the test is "A and B"), and it then resolves B in opposite fashion. Because the stated resolution of A does not affect the resolution of the test, A is a nonsupportive proposition and thus presumptively dicta. Whether A should nonetheless count as a holding may depend on the degree to which the test is ordered.	Justice Powell applied the strict scrutiny test, which may be understood to require first a compelling interest prong, and second means that are narrowly tailored in furtherance of that interest. The Court applied the test in that order, concluding that diversity was a compelling interest, but that the Davis plan was not narrowly tailored to it. The question is whether the compelling interest conclusion should be considered a holding because it is the first prong of the strict scrutiny test.

III. POSITIVE MODELS AND NORMATIVE FACTORS

We have now identified several categories of judicial assertions that courts can, and sometimes do, present as holdings in their judicial opinions. Before reviewing the various definitions of holding and dicta that have been offered in the literature and presenting our own, the subject of Part IV, we must lay a proper theoretical foundation to use in evaluating the various definitions. This task has both a positive and a normative component.

Understanding the conceptual origins and characteristics of stare decisis, and defining holding and dicta, are distinct but overlapping projects. Each set of inquiries focuses in significant part on evaluating two aspects of judicial overreaching, first the risk that judges will present overly broad written opinions as precedent, and second, the risk that judges will disingenuously construe past precedents to avoid results with which they disagree. As a result, it is important to our project to understand the role of both sets of legal constraints—stare decisis and the definitions of holding and dicta—in shaping appropriate judicial behavior. This task requires that we explore the conceptual foundations and characteristics of stare decisis, and that we evaluate the extent to which stare decisis itself inhibits the two identified forms of judicial overreaching. Once we understand the constraining role of stare decisis, we can better appreciate any remaining opportunities for these forms of overreaching even with stare decisis in place. Proper definitions of holding and dicta should be targeted toward reducing (even if they cannot altogether eliminate) the remaining opportunities for inappropriate judicial behavior.

As we will demonstrate, stare decisis doctrine provides meaningful limits on opportunities on judicial overreaching. Our analysis reveals that stare decisis encourages a process of judicial decision making in which holdings are motivated by the need to resolve cases rather than by the desire to further judicial preferences concerning legal policy. It does so by encouraging judges to rely upon a set of material facts that emerge from the case before them in their effort to explain the basis for their selected judgment. As we have already seen however, the concept of materiality involves a substantial element of choice and is therefore neither self-defining nor self-limiting. Thus, while stare decisis furthers proper judicial behavior, it leaves substantial room for differences among judges, even when those judges are acting in good faith, respecting the permissible extent to which a court can properly claim holding status in a given opinion for particular propositions. It also leaves substantial room for differences as to which a

later court can credibly claim a basis for distinguishing a seemingly problematic precedent.¹²⁰ While stare decisis provides a core set of principles for limiting judicial overreaching, it does not answer several of the specific questions identified in Part II concerning those holding categories that prove the most difficult, and in fact underscores the need for a shared understanding of holding and dicta.

To understand the nature of the limits that stare decisis imposes, it is important to appreciate the doctrine's theoretical origins. As we will elaborate,¹²¹ prior commentators have offered models grounded in rational choice theory to demonstrate that even without formally imposed rules demanding adherence to precedent, judicial systems would develop a legal regime that employs precedent in some form.¹²² In addition, commentators have used rational choice to demonstrate that within a legal system based upon precedent, judges behaving rationally are motivated to limit, to a considerable extent, the scope of their written opinions. Rational choice analysis is appealing in this context because it suggests that even if society did not impose formal obligations on judges to adhere to precedent, or to limit the scope of their opinions in some way, a system resembling the one that we observe, which possesses these characteristic features, would likely emerge as a product of the rational expectations of interactive jurists.

While these conclusions are largely compatible with our own analysis, we are less persuaded that the models developed thus far accurately reflect the dynamics that give rise to the stare decisis doctrine or to the incentives in shaping the judicial writing and construction of opinions. Exploring the contributions and limits of existing models, and suggesting some modest refinements, will allow us to appreciate several critical features of stare decisis that are of particular relevance to our larger project, namely devising theoretically sound and functional definitions of holding and dicta that complement stare decisis in encouraging appropriate judicial behavior. Refining the existing models allows us to improve upon the explanatory force of rational choice in providing a theoretical foundation for several important aspects of judicial

¹²⁰ See *supra* text accompanying note 103 (discussing the relationship between the holding-dicta distinction and the task of distinguishing precedents).

¹²¹ See *infra* Part III.A.

¹²² One datum strongly supportive of this intuition is that civilian regimes go further than declining to impose an obligation of precedent, and instead formally eschew precedent. See JOHN H. MERRYMAN, *THE CIVIL LAW TRADITION: AN INTRODUCTION TO THE LEGAL SYSTEMS OF WESTERN EUROPE AND LATIN AMERICA* 144 (1969). While civilian regimes formally reject precedent, legal scholars have observed that over time, such systems nonetheless begin to resemble common law systems in their effective, if not formal, reliance upon precedent. See *id.* (observing that “although there is no formal rule of *stare decisis*, the practice is for judges to be influenced by prior decisions”); *infra* note 265.

practice that we commonly observe. Once we have accurately described the nature and limits of stare decisis based upon our rational choice model, and have further considered the limits of that doctrine in properly constraining the two feared manifestations of judicial overreaching that we have identified,¹²³ we can then evaluate the role of proper understandings of holding and dicta in furthering appropriate judicial incentives. After we set out our positive model, we will thus undertake a careful analysis of the various categories of judicial assertions set out in Part II based upon four normative considerations that are particularly relevant to our larger project.

Before proceeding, we would like to offer a few brief methodological comments. Our analysis is not intended to suggest (and does not suggest) that judges uniformly exhibit a set of common behaviors that one can label “rational.” We do not, for example, suggest a single maximand or utility function that can be superimposed on judges as a class.¹²⁴ Judges, like members of any other profession, have a wide range of personal attributes and expectations, including expectations about their profession. Our goal is not to use rational choice to describe the psychology of the individual jurist or even of jurists generally. Instead, by developing some basic suppositions about judges and recognizing that there are exceptions to any characterization that we make, we can evaluate how the judiciary as an institution is likely to respond to some of the problematic phenomena that our analysis reveals. We can then further consider how judges behaving rationally are likely affected by those judicial practices that our analysis predicts will emerge. The essential intuition that underlies our analysis is merely that institutions are formed or modified in response to particular difficulties that confront individuals when they are unable to solve certain bargaining or coordination problems on their own, and further, that individuals shape their behavior, and their expectations, in consequence of the structures of the institutions in which they are operating.

Our analysis will demonstrate that solutions to the problems that we identify sometimes affect the structure of the judiciary as a whole; sometimes take the form of rules (or rebuttable presumptions); and sometimes, when changes in institutions or rules prove impracticable, take

¹²³ See *infra* Part III.A.

¹²⁴ To that extent, we agree with the subtitle of Judge Posner’s essay, responding to the question “What do judges maximize?” with the reply “the same thing as everyone else.” See Richard A. Posner, *What Do Judges Maximize? (The Same Thing as Everybody Else Does)*, 3 S. CT. ECON. REV. 1 (1993). The point of course is that individual preferences—and in this case judicial preferences—are widely varied. A sound rational choice analysis can account for such differences and still provide insights into how the judiciary shapes, and is shaped by, the expectations of a group of diverse judges acting rationally in pursuit of their objectives.

the form or customs or norms.¹²⁵ While the solutions to the problems that we identify vary, in each instance, rational choice analysis is helpful in identifying the nature of the problem, the likely judicial responses, and the probable solution. To be clear, we do not contend that any of the solutions that we identify are perfect in the sense of absolutely ensuring appropriate judicial incentives and behaviors. Indeed, our project is motivated by the limits of existing institutional structures, rules, and norms in encouraging proper judicial behavior in the writing and construction of opinions, and specifically by the need for improved definitions of holding and dicta.

Before proceeding to our analysis of the models, we wish to acknowledge that rational choice has been criticized for relying upon thinly constructed models to justify bold public policy prescriptions.¹²⁶ Indeed, we are sympathetic to such concerns.¹²⁷ Thankfully, our purpose in assessing and developing rational choice models of stare decisis here is quite modest. We are using rational choice to help answer some of the fundamental questions about our system of precedent as it actually exists. To the extent that our model helps to answer such questions, and to the extent that our answers help to provide an important context for constructing proper definitions of holding and dicta, this part of our project has served its purpose. We are not using rational choice to suggest an optimal stare decisis rule, which is beyond the scope of our project, or to suggest how holding and dicta should be defined. For that, one must undertake a careful normative analysis, which although edified by our positive analysis, is the subject of the next subpart.

Instead, the rational choice analysis we offer is targeted to providing answers to the following inquiries: Why would judges willingly limit their ability to effectuate legal policy by abiding the precedents of other judges with whom they disagree?¹²⁸ If judges on the same level

¹²⁵ For an interesting recent article describing the role of stare decisis as a norm constraining decision making on the United States Court of Appeals for the Sixth Circuit, see Emery G. Lee III, *Horizontal Stare Decisis on the U.S. Court of Appeals for the Sixth Circuit*, 92 KY. L.J. 767 (2004). While we agree that stare decisis can sometimes be understood as a norm, our rational choice analysis is designed to study stare decisis within the context of several features that limit judicial overreaching, some affecting institutional structure, some taking the form of rules, and some taking the form of norms. See *infra* Part III.A.

¹²⁶ See, e.g., DONALD P. GREEN & IAN SHAPIRO, *PATHOLOGIES OF RATIONAL CHOICE THEORY: A CRITIQUE OF APPLICATIONS IN POLITICAL SCIENCE* 11 (1994) (arguing that “too often prescriptive conclusions . . . are floated upon empirically dubious rational choice hypotheses”).

¹²⁷ See, e.g., Maxwell L. Stearns, *Rational Choice and the Rational Actor*, in *ENCYCLOPEDIA OF LAW AND SOCIETY* (David S. Clark ed., forthcoming 2004) (urging rational choice theorists to develop “narrower claims increasingly subject to meaningful falsification”).

¹²⁸ Addressing this question does not require that we assume judges are bent on imposing their normative views into law. A judge’s view of legal policy can consist of a desire to produce results that are driven by a dispassionate application of external

court have an incentive to disregard precedents of other judges, how does the judiciary develop a system of precedent? Why are multi-member appellate court panels better equipped than single-judge trial courts at encouraging (or presumptively requiring) adherence to the precedents established by earlier decision makers on the same level court? Finally, and perhaps most importantly, if we assume that judges are rationally motivated to construct a system that adheres to some form of precedent, what then prevents an opinion writer from using any given case to announce the judge's preferred resolution of any number of legal policy issues and from claiming holding status for the resulting discussion? While our rational choice model will provide answers to these, and other related,¹²⁹ questions, we do not suggest that ours is the only correct or edifying approach, or that we have constructed the only plausible account of stare decisis.¹³⁰ Rather, we contend that our model offers important insights into the nature and problem of precedent that provide an essential backdrop in devising a meaningful definition of holding and dicta.

A. Modeling Stare Decisis and the Holding-Dicta Distinction

The rational choice model stems from a puzzle concerning judicial incentives in crafting opinions. Academic commentators within law and political science routinely assume that judges are motivated by the desire to implement their preferred resolutions of issues of legal policy.¹³¹ Such a claim can be advanced in a broad form, insisting that judges seek to convert desired policy objectives into law without regard to the legal issues presented, or (more plausibly in our view) in a narrow form, maintaining that when legal doctrines admit of some flexibility in their

principals of law, even those with which the judge disagrees. See MAXWELL L. STEARNS, PUBLIC CHOICE AND PUBLIC LAW: READINGS AND COMMENTARY 540 (1997) (explaining that the rational choice assumption that judges seek to impose their own normative views into law can include judges whose normative views “require principled, rather than preference-based, decisionmaking”). But a system of precedent can, nonetheless, prevent a judge from achieving even that objective if the decisions of judges who prefer to impose their normative views onto the law are binding precedent.

¹²⁹ For example, if an earlier court has issued a judicial decision that is not overreaching, what is to prevent a later court nonetheless from disregarding it to the extent that the later court finds the constraint it imposes problematic? See *infra* Part III.A.3.

¹³⁰ For two interesting alternative economic perspectives on stare decisis, see Lewis A. Kornhauser, *An Economic Perspective on Stare Decisis*, 65 CHI.-KENT L. REV. 63 (1989), which uses game theoretic analysis to evaluate costs and benefits of stare decisis when changes in technology outpace doctrinal preferences for less valuable activities; and Jonathan R. Macey, *The Internal and External Costs and Benefits of Stare Decisis*, 65 CHI.-KENT L. REV. 93 (1989), which explains the adaptability of stare decisis to societal change and discussing the role of legislation as an external check on precedent that entrenches outmoded technologies. While these two articles provide valuable assessments of the societal costs and benefits of stare decisis, they do not focus on the capacity of stare decisis to emerge as an endogenous feature of judicial decision making.

¹³¹ The classic work is JEFFREY A. SEGAL & HAROLD J. SPAETH, *THE SUPREME COURT AND THE ATTITUDINAL MODEL* (1993). We, however, are not claiming that judges necessarily seek to impose their normative views into law. See *supra* note 128 (explaining that judges can hold the view that the judicial role should be limited to the dispassionate and neutral resolution of legal disputes).

construction, judicial policy preferences will inform which of the plausible constructions a judge is likely to prefer. Either way, within our post-realist jurisprudential world, it is commonly appreciated that judges not only make law,¹³² but also that they seek to advance their views of legal policy in doing so.¹³³ Indeed, the prevalence of this view is so striking that those who embrace a contrary vision of judicial law making—one that anticipates that judges will be motivated by a neutral desire to resolve cases within the framework of existing rules and without regard to any policy preferences that they might hold—are depicted as naïve.¹³⁴ Even so, our understanding of rationality and the model we build upon that assumption is sufficiently accommodating to include judges for whom this is a sound description.¹³⁵

If judges truly are motivated by the desire to advance their views of legal policy in the cases that they decide, they have a potentially strong incentive to use those cases that appear before them as vehicles with which to make substantially broader pronouncements of legal policy than is required to resolve the immediate case. The puzzle is that while judges sometimes might be seen as overreaching, many, perhaps most, judges are cautious about creating new law, and very few seem to use cases to resolve altogether unrelated issues. The immediate challenge is to identify those mechanisms within the judicial process that help to counteract the tendency to overreach, even if they do not succeed in eliminating it entirely. This puzzle is not new. Indeed, it has motivated legal scholars in their efforts to explore the conceptual origins of both horizontal and vertical stare decisis, as well as in their efforts to explain judicial incentives to craft narrow rather than broad holdings.

¹³² As one commentator has aptly noted, “Post-realist jurisprudence must depart from the truism that judges make law and begin instead with the question of how they make law.” See Neil Duxbury, *Faith in Reason: The Process Tradition in American Jurisprudence*, 15 CARDOZO L. REV. 601 636 (1993).

¹³³ We do not intend to suggest that judges create policy in the same manner as do political actors operating within elected branches of government. Indeed, our project is intended to underscore an essential difference between adjudicatory and political decision making, namely the greater constraint—imposed by the fortuitous circumstances that give rise to a case—on judges in their efforts to create desired policy relative to legislators that a proper understanding of dicta and holding imposes. See *infra* Part III.B (describing normative considerations that emerge as a function of judicial lawmaking).

¹³⁴ Cf. Nancy Weston, *The Fate, Violence, and Rhetoric of Contemporary Legal Thought: Reflections on The Amherst Series, the Loss of Truth, and the Law*, 22 LAW & SOC. INQUIRY 733 (1997) (asserting that “[l]egal formalism . . . appears ripe for discard as naïve and illusory—or, worse, as mendacious, the later version of this judgment, as indeed it must in a world in which realism reigns”).

¹³⁵ See *supra* note 128.

1. *Modeling Horizontal Stare Decisis*

As is generally the case with rational choice analysis, we will begin with a set of stark assumptions, and then relax the assumptions to bring the model closer to the actual institutions that our analysis is offered to study. In this case, we will begin by imagining that our judiciary consists solely of a single level trial court in which individual judges hear and resolve disputes. We can imagine the judges on such a court realizing that while they would prefer that their colleagues on the bench respect their resolutions of issues they resolve as a necessary part of the cases they decide as precedent, their colleagues do not have a strong incentive to do so. If other judges agree in principle to the resolution of such issues, they will, of course, reach similar results in future cases. But this signifies no constraint in the form of precedent. After all, even absent a prior case that turned on the same issue of law presented in a new case, the later deciding judge would achieve the same result.¹³⁶

This insight can be cast in simple game theoretical terms using the framework of the standard prisoners' dilemma. While each judge would prefer to see his or her views reflected into law, each would also prefer not to be bound by the views of other judges who happened upon the need to resolve the same issue in an earlier case. As a result, without regard to whether the other judge abides one's own opinions as precedent, one has an incentive to decline to honor the other judge's opinions as precedent. These incentives are mutual and thus the dominant outcome is mutual defection even though the payoffs to each judge would be greater if each was able to achieve a regime of mutual cooperation, meaning that each respects the opinions of the other judge as precedent. This is so even though mutual cooperation itself limits the ability to achieve results that each judge might otherwise prefer in individual cases.

¹³⁶ This point has long been understood:

If a court follows a previous decision, because a revered master has uttered it, because it is the right decision, because it is logical, because it is just, because it accords with the weight of authority, because it has been generally accepted and acted on, because it secures a beneficial result to the community, that is not an application of *stare decisis*. To make the act such an application, the previous decision must be followed because it is a previous decision and for no other reason.

Max Radin, *Case Law and Stare Decisis: Concerning Prajudizienrecht in Amerika*, 33 COLUM. L. REV. 199, 200 (1933). It is possible, for example, that a prior decision explored the legal issue in a sufficiently persuasive manner that a later judge, who initially would have resolved the issue differently, changes positions. Should this occur, the second judge is not basing the resolution on the fact of the precedent, however, but rather on the persuasiveness of the earlier reasoning. A judge is invariably free to do this, just as a judge may find a treatise, law review article, an op ed, or a speech by a law professor persuasive on a relevant point of law.

We see traces of such a regime in which decision makers resolve cases but do not create precedents in doing so within some real world institutions for the resolution of legal disputes. These include civilian courts, which formally eschew precedent,¹³⁷ and both nonbinding mediation and arbitration, which do not develop binding bodies of internal precedent in the same manner as do courts.¹³⁸ While there are certainly differences between the hypothetical regime that we imagine and any of these real world regimes, for simplicity, we will refer to our decision makers for now as arbitrators. In our hypothetical regime, the individual arbitrators resolve disputes without developing precedent that binds other arbitrators. Their immediate concern is to identifying mechanisms that will enable them to elevate themselves from merely resolving disputes to judging, and in doing so, to creating, and not merely applying, law.

Professor Erin O'Hara has developed a game theoretic model, which although designed to explain *stare decisis*, ultimately demonstrates the conceptual difficulty that individual judges on the same level court confront in their effort to construct a system of precedent that elevates themselves from *de facto* mediators, who resolve disputes without making law, to judges, who perform both functions.¹³⁹ O'Hara begins with the observation that different judges sometimes hold different normative views.¹⁴⁰ As a result of this divergence of viewpoint, individual judges would prefer their colleagues to abide the decisions they issue as precedent, while, at the same time, avoiding any reciprocal obligation to abide the opinions of their colleagues as precedent.

Thus cast, O'Hara's model represents the classic prisoner's dilemma that we have identified. For each judge in the game, the ideal situation is to impose an obligation on other judges, but not to have an obligation imposed on one's self. If other judges honor precedent, it is

¹³⁷ See MERRYMAN, *supra* note 122, at 144.

¹³⁸ Arbitrators are not entirely bound by externally-imposed rules of law, whether the product of statutes, administrative regulations, or the decisions of common law courts, but are generally expected to conduct arbitration in the shadow of the law. See Daniel Q. Posin, *Mediating International Business Disputes*, 9 FORDHAM J. CORP. & FIN. L. 449, 464 (2004) (positing that "external standards [and] bargaining in the shadow of the law . . . can eliminate extreme 'outlier' possibilities and narrow the scope of negotiation/mediation"). A principal objective of mediation is to achieve predictable, and thus defensible, outcomes at lower cost than would be achieved through more costly litigation. See Alison E. Gerenscer, *Family Mediation: Screening for Domestic Abuse*, 23 FLA. ST. U.L. REV. 43, 47 (1995) (asserting that "[t]he goals of mediation include: reducing the court's docket, reducing the demand on judicial resources, accelerating the rate of case resolution, reducing the cost of resolving conflicts, increasing the litigants' satisfaction with the court system, and improving relationships between disputing parties").

¹³⁹ Erin O'Hara, *Social Constraint or Implicit Collusion? Toward a Game Theoretical Analysis of Stare Decisis*, 24 SETON HALL L. REV. 736 (1993). While O'Hara imagines single panel appellate judges, we adapt her analysis to imagine single-panel arbitrators.

¹⁴⁰ See *id.* at 743 ("[I]f Ann and Bert shared perfectly identical normative views, then the legal system would function consistently with a rule of *stare decisis*. . . . Where their normative views differ . . . , however, the judges will engage in nonproductive competition.").

rational for a judge to receive the benefit of their adherence to the judges' opinions, while declining to honor their opinions. And if other judges decline to honor the judge's precedents, it remains rational to decline to honor their precedents. Either way, it is rational to "defect" from a regime of precedent enforcement. Because these payoffs are reciprocal, meaning that other judges have the same incentives, mutual defection—meaning the absence of a system of precedent—emerges as the dominant strategy. While O'Hara describes the resulting regime, one that threatens to undermine the optimistic claim of endogenously created stare decisis, as one of "nonproductive competition,"¹⁴¹ we can also think of the resulting regime as that of arbitrators failing in their attempt to change their status to judges.

O'Hara considers whether this regime of mutual defection could be avoided if the game were treated as iterated, rather than single period. O'Hara observes that if the iterations are infinite, then the judges might well achieve a cooperative solution: "As long as the judges expect to work together indefinitely, then each can use potential future benefits from stare decisis to encourage present cooperation from the other."¹⁴² A judge who fails to cooperate in one period invites retaliation in the next period, as the judge whose decision was ignored in turn refuses to enforce the other judge's precedent.¹⁴³ "As long as the gain from defecting today is less than or equal to the present value of future losses from perpetual noncooperation, then both judges are willing to follow each other's precedents."¹⁴⁴

Professor O'Hara recognizes, however, that the cooperative solution induced by the iterated games threatens to unravel when the participants confront a certain, or anticipated, end period. To illustrate, O'Hara posits three known periods, and considers what happens in the third

¹⁴¹ *Id.* at 748-49.

¹⁴² *Id.* at 751.

¹⁴³ *Id.* at 752 ("[A]t least in this grossly oversimplified model, the judges can maintain an equilibrium of cooperation with the threat of perpetual defection."). The analysis is not intended to suggest that stare decisis is the product of endless bilateral negotiations or of endless tit-for-tat-games. See generally ROBERT AXELROD, *THE EVOLUTION OF COOPERATION* 27-54 (1984) (modeling tit-for-tat strategy). Instead, the analysis is intended to suggest the conditions under which stare decisis does, or does not, emerge as a dominant strategy. Once the regime is established, because stare decisis takes the form of a rule, even the opinions of a noncooperative judge will be respected as precedent. Rather than punishing through disregarding a noncooperative judge's precedent, the defecting judge will suffer as few other judges will join that judge's opinions. And since stare decisis, as shown below, see *infra* notes x-x and accompanying text, emerges in the first instance within multimember appellate panels, this is a potentially effective sanction against defectors, barring conditions in which two judges agree to disregard precedent. Even then, en banc review remains an additional check.

¹⁴⁴ *Id.* Rather than "solving" a prisoners' dilemma, this "solution" avoids one by altering the payoffs for cooperative and noncooperative behavior such that the regime no longer represents a prisoners' dilemma. See STEARNS, *supra* note 128, at 541-42. Eric Rasmusen, in a separate article written at about the same time as O'Hara's, see Eric Rasmusen, *Judicial Legitimacy as a Repeated Game*, 10 J.L. ECON. & ORG. 63 (1994), reaches a similar conclusion concerning the possibility of a cooperative solution under specified conditions.

period when the players anticipate no further iterations.¹⁴⁵ In this situation, the players no longer confront a threat in any future period for defection, and, as a consequence, mutual defection reemerges the dominant strategy as in a single period game. In period 2, anticipating mutual defection in period 3, the players again defect, realizing that they will receive no benefit from cooperation. With any definite end period, the same “unraveling” begins in the final period and continues to the very first period, suggesting that unless iterations are infinite, or sufficiently long that future defection is heavily discounted, unraveling threatens to restore mutual defection as the dominant outcome.¹⁴⁶

We do not intend to suggest that the choice is quite as stark as selecting between unraveling with anything short of endless iterations versus cooperation in an infinitely iterated game. Some literature suggests that even without endless iterations, under certain assumptions it is possible to generate cooperative solutions.¹⁴⁷ It is beyond our scope here to review the literature on the iterated prisoners’ dilemma and assess the possibility of cooperation where there is only some probability of the game ending in any period. Rather, it is sufficient to note that a cooperative solution that generates a regime of *stare decisis* among single judge panels on the same level court is likely to be at least substantially unstable. And in fact, this intuition, which is consistent with O’Hara’s analysis, reflects practices that we observe in trial courts.¹⁴⁸

Within the United States, as a general matter, precedent does not take as strong form a constraint—if it is a constraint at all—among trial judges operating in the same level court. While judges might find the opinions of their colleagues persuasive, they do not generally consider themselves bound to adhere to those opinions as precedents. This suggests that the

¹⁴⁵ O’Hara, *supra* note x, at 774-75 n.72.

¹⁴⁶ For helpful discussions of the prisoners’ dilemma and the associated difficulty of unraveling in repeated games, see ROBERT COOTER & TOM ULEN, *LAW AND ECONOMICS* 216-17 (3d ed. 2000); and ERIC RASMUSEN, *AN INTRODUCTION TO GAME THEORY* 88-89 (1989).

¹⁴⁷ See Ronald J. Gilson & Robert H. Mnookin, *Disputing Through Agents: Cooperation and Conflict Between Lawyers in Litigation*, 94 COLUM. L. REV. 509, 520 (1994) (positing that depending upon the payoff function, cooperative behavior can arise in a multi-round prisoners’ dilemma when players anticipate a high possibility of a subsequent round in which cooperation is rewarded and defection is punished.); David M. Kreps et al., *Rational Cooperation in the Finitely Repeated Prisoners’ Dilemma*, 27 J. ECON. THEORY 245, 245-47 (1982) (explaining conditions under which cooperation may arise even in finite-period repeated games).

¹⁴⁸ Thus, Professor O’Hara notes that “[i]t is unclear whether district courts actually follow a rule of horizontal *stare decisis*.” *Id.* at 773 n.66. Similarly, in Charles H. Nalls & Paul R. Bardos, *Stare Decisis and the U.S. Court of International Trade: Two Case Studies of a Perennial Issue*, 14 FORDHAM INT’L L.J. 139 (1990-91), the authors observe: “Although most treaties consider a district court judge bound by the decisions of his colleagues on the court, case law shows that some judges do not believe this to be the case. Instead, they view the precedential effect of other judges’ decisions as persuasive, but not binding authority.” *Id.* at 146.

practice thus far tracks the observations of the model, which would not predict a strong form cooperative equilibrium in this context. And yet, the practical problem remains, namely that individual trial judges seek to extricate themselves from the function of arbitrating (or merely resolving disputes) to that of judging (thus making law as needed in the course of resolving disputes). The solution to this problem lies not in the iterated prisoners' dilemma, but rather in the nature of a pyramidically structured court.

2. *Modeling Vertical Stare Decisis*

In an earlier article,¹⁴⁹ Professor William Landes and then-Professor Richard Posner also addressed the conceptual difficulty in employing a prisoners' dilemma analysis to devise an explanation for horizontal stare decisis. Landes and Posner offered an alternative explanation for the emergence of stare decisis doctrine that relies upon the emergence of pyramidically structured judicial system.¹⁵⁰ By devising another level in an emerging judicial pyramid, judges create a means of developing and promoting adherence to precedent. Most importantly, in doing so, they avoid the need for cooperation among the judges at the trial court level.¹⁵¹ An appeals court's "power to reverse the decisions of lower courts," Landes and Posner explain, "checks any tendencies on the part of lower-court judges to disregard precedent (reversal foils a judge's attempt to create his own precedent), and its own position in the judicial hierarchy checks its members' tendencies in that direction."¹⁵² An appellate court, even one that is drawn from the ranks of the trial court judges, most logically in odd-numbered random groupings to prevent ties, is able to induce cooperation and limit defection among individual trial judges who would otherwise fail to make law as they routinely disregard each others' precedents. The analysis reveals that a statutory mandate or other external source of law is not required to induce cooperation in the form of precedent because the emerging judicial hierarchy, which creates the same result, emerges as a function of the rational incentives of participating jurists.

¹⁴⁹ See William M. Landes & Richard A. Posner, *Legal Precedent: A Theoretical and Empirical Analysis*, 19 J.L. & ECON. 249 (1976).

¹⁵⁰ Civilian regimes, which formally eschew precedent, also follow a pyramidic structure. This does not undermine the Landes and Posner theory, but might be attributable to the need to correct rulings within the immediate case, rather than to impose the obligation of precedent, or it might be due to the de facto practice of following precedent—without doing so formally—within civilian courts. See *supra* note 122; *infra* note 265.

¹⁵¹ As explained below, their proffered solution does not actually induce mutual cooperation because it does not result in adherence to precedents issued at the trial court level. Instead, it solves the problem by imposing a vertical restraint.

¹⁵² Landes & Posner, *supra* note 149, at 273.

While this analysis helps to resolve the difficulty posed by mutual defection among judges at the trial court level, the explanation remains only partial. The same incentive to ignore precedent that confronts individual trial judges also confronts the various members of separate appeals court panels. The question then arises why such panels do not experience the same mutual defection—or nonproductive competition—that trial judges would experience in a judiciary that lacked a hierarchical structure. To some extent this motivation is tempered when—as occurs within the United States Courts of Appeals—appellate panels include three judges who are randomly drawn. For an appellate judge to disregard precedent, the appellate judge will have to find at least one other judge who is willing to sign on. In addition, anticipating serving on the same panel with a judge whose opinion one chose to disregard might motivate more collegial behavior. But a considerable incentive to disregard precedent, perhaps by engaging in disingenuous distinctions from prior rulings, remains, especially where judges on a panel have similar ideological preferences.¹⁵³ The incentive might even be exacerbated at the appellate level as a result of the greater impact of precedent over a broader number of future trial and appellate judges.

A final step in the judicial pyramid, namely the creation of a supreme court, to the extent that it serves a similar function with respect to appeals courts that appeals courts serve with respect to district courts, might further limit incentives to disregard precedent. But at best such a solution once again is only partial.¹⁵⁴ Within the federal judiciary, as is widely recognized, Supreme Court review is sufficiently rare—far rarer than appeals court review of district court decisions¹⁵⁵—that substantial opportunities remain to disregard prior appeals courts decisions as

¹⁵³ This incentive has led some commentators to argue that appellate courts should self-consciously seek to include broad political representation in panels. See Emerson H. Tiller & Frank B. Cross, *A Modest Proposal for Improving American Justice*, 99 COLUM. L. REV. 215, 220-24 (1999) (proposing that panel assignments ensure split-party representation to reduce probability of partisan influence in federal appellate decision making); see also Michael Hasday, *Ending the Reign of Slot Machine Justice*, 57 N.Y.U. ANN. SURV. AM. L. 291 (2000) (refining the Tiller-Cross proposal by proposing a mechanism in which litigants provide preference-orderings of appellate judges).

¹⁵⁴ We might imagine some concerns for defection within a single appeals court remaining, and the use of an en banc practice—calling on all active members from which randomly drawn panels are selected—to check against such defection. The Supreme Court presumably cares less about intra-appeals court defection than about adherence among appellate court panels to its precedents, so en banc review and Supreme Court review serve complementary functions. Cf. Michael Abramowicz, *En Banc Revisited*, 100 COLUM. L. REV. 1600 (2000) (arguing that en banc review could be restructured to ensure greater conformity among the circuit courts to Supreme Court precedent).

¹⁵⁵ For relevant statistics, see JONATHAN MATTHEW COHEN, APPELLATE COURTS INSIDE AND OUT: THE IMPACT OF COURT ORGANIZATION ON JUDICIAL DECISION MAKING IN THE UNITED STATES COURTS OF APPEALS (Mich. 2002), at 7 tbl.1 (“Appeals Commenced, Terminated, and Pending in the U.S. Courts of Appeals”), 43 tbl.2 (“Appeals Terminated in the U.S. Courts of Appeals and Petitions for Certiorari Filed, Granted, and Decided in the U.S. Supreme Court”).

precedents. Moreover, and perhaps more importantly, while en banc review (the practice whereby the full circuit court has the power to vacate and review three judge panel decisions) might be used to correct defection taking the form of failures to abide internal circuit court precedent, it too is also rarely used.¹⁵⁶ We know of no cases in which the Supreme Court has reviewed a case to ensure consistency of case law *within* a particular circuit. In addition, once the Supreme Court is established, it is potentially subject to a similar form of defection, here an incentive among individual justices to disregard the Court's own precedents, assuming they can persuade the requisite majority to join.

If each level in the judicial hierarchy substantially reduces the risk of defection in the court below, but cannot prevent defection at the newly created court level, the question thus arises how the various appeals courts—including the circuit courts and the Supreme Court—overcome the incentives of their own members to defect from a cooperative stare decisis regime. The answer does not rest in the prisoners' dilemma analysis, but rather lies in the need to stabilize doctrine in multimember appellate courts. While the need to stabilize doctrine is a general problem on multimember courts, even those that hear case en banc, the problem is particularly acute in courts in which subgroups, for example panels of three, hear cases for the entire court, but might not reflect the views of the court as a whole. An analysis based upon social choice demonstrates the nature of the problem of doctrinal instability on multimember appellate courts, whether or not the courts hear cases in panels, and the function of stare decisis in horizontal form in substantially reducing it.

Under specified conditions, judges on a multimember courts sometimes hold preferences over the resolution of cases,¹⁵⁷ that when subject to a series of direct pairwise contests would result in the phenomenon of cycling. For a simple illustration, consider three persons—P1, P2, P3—holding the following rank ordinally ranked orderings over options ABC: P1: ABC; P2: BCA; P3: CAB. If the three people were presented with the three possible binary comparisons and voted sincerely based upon their expressed ordinal rankings, the result would be a cycle, such that B is preferred to C (P1 and P2 winning); and A is preferred to B (P1 and P3 winning),

¹⁵⁶ See Arthur D. Hellman, *Precedent, Predictability, and Federal Appellate Structure*, 60 U. PITT. L. REV. 1029, 1049 (1999) (“En banc review is a rare occurrence.”).

¹⁵⁷ For a similar analysis of preferences over issues within cases, see STEARNS, *supra* note x, at 97-156 (illustrating with *Miller v. Albright*, 523 U.S. 420 (1998), and collecting additional cases).

implying that A is preferred to C, when in fact, C is preferred to A (P2 and P3 winning). Thus, even though each member holds preferences that are transitive, the group as a whole prefers A to B to C to A. Social choice theory demonstrates that a general assumption of individual rationality, namely that preferences satisfy the minimal condition of transitivity, cannot be assumed for groups of three or more.¹⁵⁸ With a set number of options represented by n (here three), a regime that allows only n minus 1 (here two) votes provides a stable outcome, albeit one that turns on the order in which the votes are taken. The following example will demonstrate that stare decisis in horizontal form effectively removes an option for multimember appellate courts, thus stabilizing doctrine, with the effect of rendering doctrine dependent on the order in which cases are presented.

Imagine a three judge panel deciding two cases on the same day,¹⁵⁹ first, *Adams v. School Board*, raising a constitutional challenge to busing ordered by a school board, and the second, *Barnes v. State Court*, raising a constitutional challenge to busing ordered by a state court. Imagine that in each case, the lower court rejected the constitutional challenge and that the same three judge panel—consisting of judges J1, J2, and J3—decided the two cases on the same day. In *Adams*, J1 and J2 conclude that the challenge is not meritorious, and thus affirm the decision below sustaining the statute, while J3 dissents. In *Barnes*, J2 and J3 conclude that the challenge to the state court busing order is meritorious, and thus reverse the lower court decision sustaining that order, while J1 dissents. Imagine that in their respective dissenting opinions, J3 and J1 both state that in their views, the two cases are indistinguishable, a position that is consistent with their individual votes to either reject (as did J1) or sustain (as did J3) both challenges. The analysis reveals three majorities, one of which must be suppressed: (1) J1 and J3 to sustain the statute in *Adams*; (2) J2 and J3 to strike the court order in *Barnes*; and (3) J1 and J3 to treat *Adams* and *Barnes* as indistinguishable.

¹⁵⁹ This discussion is based upon a stylized rendition of two actual Supreme Court cases. For a more detailed discussion and analysis, see *id.* at 170-77, which illustrates intransitive Supreme Court preferences using *Crawford v. Board of Education*, 458 U.S. 527 (1982), and *Washington v. Seattle School District No. 1*, 458 U.S. 457 (1982), both of which were issued on the same day.

Deleted: ¶

Now imagine that instead of the two cases arising concurrently, *Adams* arises first, followed, one year later, by *Barnes*. Because the *Adams* court affirms (J1 and J2 controlling), and because J1 and J3 believe that the cases are indistinguishable, the result in *Barnes* would also be to affirm. Alternatively, if *Barnes* arises first, resulting in a reversal (J2 and J3 controlling), then because J1 and J3 view the cases as indistinguishable, the result in *Adams* will also be to reverse. The analysis reveals that depending upon the order of presentation, any of the three identified outcomes might arise: both upheld, both struck down, or *Adams* upheld and *Barnes* struck down.¹⁶⁰ In each sequence above, the court has not permitted resolution of all three issues (*Adams* as a matter of first impression; *Barnes* as a matter of first impression; and *Adams* or *Barnes* with the other case as a precedent) that would be required to determine that the combined outcomes over the three possible binary comparisons are an arbitrary function of the order in which the cases have been presented for consideration. *Adams* then *Barnes* generates opposite outcomes from *Barnes* then *Adams*, and both sequences generate a different outcome from the two cases arising together and thus without one or the other as precedent.

If the court were permitted to run through all possible iterations endlessly, the arbitrariness would become clear.¹⁶¹ Social choice theory, of course, does not posit that institutions are plagued by endless cycling and indeterminacy. In most instances, institutional structures induce equilibria even when the participants possess intransitive preferences.¹⁶² Courts are no different in this regard, and our example serves to illustrate this point. At whatever point the relevant structure induced equilibrating rule breaks the theoretical cycle, the chosen outcome inevitably suppresses a set of majority preferences for a contrary result.

This admittedly stylized presentation provides a core insight into the function that stare decisis serves in multimember appellate courts. Social choice analysis reveals that when a group

¹⁶⁰ The fourth possible outcome—*Adams* upheld, *Barnes* struck down—is not necessary to the analysis.

¹⁶¹ If the Court first decided *Adams*, it would affirm, and then, based upon the *Adams* precedent, it would also affirm in *Barnes*. But a majority favoring a reversal in *Barnes* in the absence of a governing precedent has been suppressed. The *Barnes* affirmance resulted from applying *Adams* as precedent in *Barnes* even though absent that precedent, some of the arguments that had been rejected in *Adams* might have proved sufficiently persuasive to some judges to produce a different outcome in *Barnes*. Following the *Barnes* affirmance, if those arguments rejected in *Adams* could be resurrected, a second *Barnes* case, call it *Carey*, could be reversed. If so, and if *Carey* were then treated as precedent, a second case like *Adams*, call it *Darby*, would also result in a reversal. And we could now start the ball rolling all over again. There is no stable outcome here because in a regime with endless iterations and with intransitive preferences, a theoretical majority opposes every possible outcome. See generally STEARNS, *supra* note x, at 41-96 (describing the problem of social choice and providing illustrations).

¹⁶² For the seminal work on structure-induced equilibria, see Kenneth Shepsle & Paul Weingast, *Structure-Induced Equilibrium and Legislative Choice*, 37 PUB. CHOICE 503 (1981).

possesses cyclical preferences over multiple options and must achieve determinate outcomes, one means of ensuring actual results is to limit the number of binary comparisons relative to the number of options.¹⁶³ In parliamentary procedure, for example, rules proscribing reconsideration of defeated alternatives¹⁶⁴ serve the function of breaking cyclical indeterminacy by ensuring that for n options, there will be no more than $n - 1$ binary comparisons.¹⁶⁵ The outcome under such a regime is arbitrary; had the voting started elsewhere, the outcome would have been different. Because any outcome is arbitrary, however, the regime of horizontal stare decisis has the decided benefit of giving at least the appearance of majority support, as the final outcome and all those preceding it will be decided with majority opinions, within the context of a legitimating process.

Thus, stare decisis is the judicial equivalent of a rule limiting the number of binary comparisons relative to options in appellate courts.¹⁶⁶ The doctrine achieves this by holding off limits those options that have been defeated in earlier cases, which now form the basis of precedent. The analysis demonstrates that while vertical stare decisis—the product of pyramidal courts—emerges as a solution to the problem trial judges face in trying to extricate themselves from an arbitration type regime in which they resolve disputes but do not make law, the horizontal form of stare decisis within appellate courts emerges as a solution to a different problem, namely the need ensure stable doctrine and prevent cycling. Stare decisis may not have arisen historically to solve a potential cycling problem. Regardless of its initial motivation, horizontal stare decisis has survived at least in part because the doctrine solves a practical difficulty that might otherwise threaten the objective of doctrinal stability on multimember appellate courts.

The analysis also helps to explain a difference in the nature of stare decisis as it operates at the level of the Supreme Court and in the United States Court of Appeals (and by extension between state supreme courts and intermediate state courts of appeals). In the Court of Appeals, panels are the product of random draws of three among a larger set of members of the court. The randomness exacerbates the problem of doctrinal instability by increasing the probability that

¹⁶³ See STEARNS, *supra* note 38, at 158.

¹⁶⁴ See *id.* at 70. For an important article that reviews the Rules of Parliamentary from a social choice perspective, see Saul Levmore, *Parliamentary Law, Majority Decisionmaking, and the Voting Paradox*, 74 VA. L. REV. 971 (1989).

¹⁶⁵ See STEARNS, *supra* note 38, at 170.

¹⁶⁶ See *id.* at 158. For a discussion of congressional practices that limit the number of votes relative to options, see William H. Riker, *The Paradox of Voting and Congressional Rules for Voting on Amendments*, 52 AM. POL. SCI. REV. 349, 354 (1958).

panels that do not represent accurately the membership of the court as a whole will decide cases.¹⁶⁷ In contrast, with narrow exceptions, the Supreme Court always meets as an en banc court. As a result, although that Court is occasionally prone to preference aggregation problems that conceal doctrinal cycles,¹⁶⁸ the problem is minimized relative to circuit courts because subgroups generally lack the power to resolve cases on behalf of the Court as a whole.¹⁶⁹

The Courts of Appeals have two additional mechanisms for reducing doctrinal instability. The first is the possibility of Supreme Court review, which as we have already explained is not likely motivated to check against intra-circuit doctrinal instability. The second is en banc review by the full membership of the relevant court of appeals.¹⁷⁰ While this is a statistical rarity as compared with review by the Courts of Appeals of district court cases, for purposes of stare decisis enforcement, en banc review nonetheless serves a similar function with respect to the circuit court as a whole to that which circuit court review serves for district courts. A possible justification for en banc review is therefore to check against panels that stray sufficiently from predictable outcomes for the circuit as a whole that they invite a considerable risk that over time, the result will be to codify unstable—and perhaps eventually cyclical—preferences.

Within our judicial system, we observe several features that the model outlined here would predict: (1) a relatively weak stare decisis rule among trial judges operating at the same level court; (2) a substantially stronger vertical stare decisis norm operating from a supreme court to an appellate court and from an appellate court to a trial court; and (3) a strong presumptive stare decisis rule, one that requires overruling in either a higher court or an en banc court operating horizontally at the supreme court and appeals court levels. The solutions to the

¹⁶⁷ This problem may be particularly acute in the United States Court of Appeals for the Ninth Circuit, which uses a mini-en banc procedure that allows 11 of the active 26 judge membership to review three-judge panel decisions. For a probabilistic analysis that explores the likely differences in outcomes as between full en banc review and mini-en banc review in the Ninth Circuit, see Maxwell L. Stearns, *Appellate Courts Inside and Out*, 101 MICH. L. REV. 1764, 1780-85 (2003) (reviewing COHEN, *supra* note 155).

¹⁶⁸ See generally STEARNS, *supra* note 38, at 138-56 (illustrating cycles within individual cases); *id.* at 179-90 (illustrating cycling with groups of cases over time).

¹⁶⁹ Exceptions include single justices acting as circuit justices. See generally S. CT. R. 22 (allowing for applications to individual Justices in specified circumstances). Actions by single Justices are generally not considered to have precedential value, which may explain why a comprehensive collection of such opinions were compiled and published only recently. See 1-3 IN CHAMBERS OPINIONS (Cynthia Rapp ed., 2004).

¹⁷⁰ But see *supra* note 167 (discussing the alternative regime employed by the Ninth Circuit). Even the Ninth Circuit allows full en banc review, but since the implementation of mini en banc review, the Ninth Circuit has never employed this further review option. See *id.* COHEN, *supra* note 155, at 182 n.22 (observing that as of the time of publication, “the judges on the Ninth Circuit have voted on three calls for a full-court en banc, but a majority of the judges has not yet voted to rehear a case decided by an eleven-judge panel.”).

difficulties of extricating jurists from the task of arbitrating to that of judging, however, creates a new difficulty of its own. Once the system of stare decisis is in place, what is to prevent its abuse? Specifically, if judges on appellate courts have the power not merely to resolve cases, but also to establish precedents, what then prevents them from abusing that power by resolving substantially more than is required in the immediate case in an effort to engraft their strongly held policy views into the fabric of legal doctrine?

3. *Modeling the Breadth of Holdings*

If we accept the premise that judges are motivated to advance their views of legal policy, whatever their views happen to be,¹⁷¹ a substantial risk arises that ambitious judges operating within a stare decisis system will use the cases that they decide to announce their preferred views of law and public policy as “holdings.” This difficulty is exacerbated once we realize that stare decisis takes its strongest form in vertical form and at the appellate level. Because courts of appeals decisions have a substantially broader impact—they apply to more courts and over a broader geographic domain—than do decisions of district court judges, the incentive to present expansive opinions as holdings is all the more powerful. So the question arises what, if anything, mitigates judicial incentives to overreach, and specifically, what motivates judges to limit the reach of their holdings.

Legal scholars who have used rational choice to study the emergence of stare decisis in its various forms have also employed rational choice theory to identify mechanisms that induce cooperative strategies that are conducive to a properly limited judicial decision making function, one that generally inhibits judges in their efforts to use particular cases to advance broad policy prescription under the guise of holdings. As with those models designed to construct a system of stare decisis, these related models suggest that even without an externally imposed rule demanding appropriate judicial restraint, it is rational for interactive judges to develop a system of appropriate restraint, one that we might think of as a custom or norm,¹⁷² that encourages

¹⁷¹ Strongly held policy views can include furthering the objective of a very limited judicial role. *See supra* note 128 and accompanying text.

¹⁷² We use the term “norm” to suggest solutions to games involving cooperative and noncooperative strategies for which the solution does not rest in a formalized rule. *See, e.g.,* ERIC A. POSNER, *LAW AND SOCIAL NORMS* 5 (2000) (modeling social norms as “a signaling game in which people engage in behavioral regularities in order to show that they are desirable partners in cooperative endeavors”); *see also id.* at 11-46. Here, jurists can signal cooperative behavior through the nature of the opinions that they write as a potential solution to the defection strategies that would predominate if instead they systematically used cases as vehicles for effectuating desired policy objectives unlinked to the cases before them. But because there is no operational rule

restricting holdings to the resolution of issues bearing a meaningful connection to the operative facts of the cases that they are called upon to decide, and thus inhibiting judicial efforts to use cases as vehicles for effectuating unrelated, or tangentially related, views of law and public policy.

In contrast with the structural aspects of stare decisis described in the last subpart, which produce enforceable rules that all participating judges are at least presumptively required to follow, here we are concerned with norms concerning the form that opinions take once the stare decisis regime is established. While those declining to adhere to stare decisis are subject to a set of formal processes, most notably the possibility of further appellate review, and difficulties in finding judges to join their opinions, there are no such formal or informal processes through which to counteract clumsily written or overbroad judicial opinions. There is no meaningful prospect for devising formal rules that appropriately tailor judicial incentives, but over time, customary practices taking the form of norms are likely to develop, and these in turn become meaningful benchmarks against which other judges on the same court can evaluate the quality of opinions and make appropriate suggestions for revision. Our analysis here and in Part III.A. reveals that rational choice explains a range of mechanisms that encourage proper judicial restraint respecting the impact and content of opinions, including institutional modifications (developing a judicial hierarchy with an imposed rule of vertical stare decisis), formal rules (creating a regime of horizontal stare decisis to induce doctrinal stability among multimember appellate panels), and norms (encouraging appropriately limited statements of holdings within opinions).

While Professor O'Hara's argument focuses on providing a game theoretical analysis of interactive judges, she also considers whether limits of judicial prescience in anticipating the consequences of overly broad holdings might provide an endogenous incentive to limit ambitious judicial opinions.¹⁷³ Thus, for example, O'Hara imagines a judge confronted with a case requiring her to determine if adult pornography is protected speech. This judge is generally protective of free speech, including pornography, but fears that an overly broad statement of the

that can insist that judicial opinions be crafted carefully, be limited to particular discussions, or omit extraneous details, any general practices that emerge concerning the scope of opinions take the form of a norm or custom, rather than of a rule.

¹⁷³ See O'Hara, *supra* note x, at 741 n.19. While this argument is not central to Professor O'Hara's larger thesis, evaluating the argument is important to our analysis because it suggests the possibility of an additional independent endogenous incentive limiting judicial overreaching.

holding might unwittingly also protect child pornography, which she does not seek to protect. The result is that the judge articulates a relatively narrow holding, perhaps one limited to the material facts, which do not involve child pornography, as a means of avoiding the pitfall of potentially codifying into law a result that does not accurately reflect her ultimate views of legal policy. O’Hara summarizes her intuition as follows: “Although the judge can later abandon that part of [the] opinion which creates undesirable incentives, [the judge] prefers to avoid encouraging such behavior in the interim.”¹⁷⁴

One difficulty this hypothetical highlights is that while some judges might be reticent to anticipate future legal issues, either based upon concerns for failing to anticipate the consequences of their holdings or simply due to their understanding of a limited judicial function, other judges might not share the same reticence. If so, then those judges who are more disposed to expand present cases into future doctrinal areas would then have an advantage in setting out the law at the expense of the law making power of more reticent judges. The result would then be a regime in which ambitious judges dominate the judicial lawmaking game. So viewed, self restraint motivated by failing to anticipate the consequences of one’s own rulings seems unlikely to promote a stable regime that limits judicial overreaching. While this might explain a tendency that we actually see, namely varying degrees of ambition in judicial opinions, it does not explain any generalized custom or norm that limits judges in general as it relates to the scope of judicial opinions.

Professor Eric Rasmusen has provided an alternative account of how judges, behaving rationally, might achieve an equilibrium in which they limit the scope of their written opinions. In his model, each judge seeks to limit the power of other judges to state overly broad holdings in exchange for other judges respecting her precedents in their own opinions. Rasmusen describes the problem, and his suggested solution, as follows:

The distinction between *ratio decedendi* and *dictum* limits the rate at which new law can be created, which is useful with regard to maintaining a precedent-obeying equilibrium. Judges must wait to create new law until appropriate cases arrive at their benches, and only the *ratio decidendi* or *holding*—the case’s essential and decisive point of law—is binding precedent. *Dicta*, rulings irrelevant to the case outcome, are, by convention, not binding on future judges, although like anyone else’s writings, they may be considered as arguments. If *dicta* were binding, a judge could create an unlimited amount of new law. His 600-page

¹⁷⁴ *Id.*

decision on a bankruptcy case could also control the law on abortion, copyright, and criminal evidence. Such an equilibrium could not be maintained . . . because even if the *dicta* carefully avoided violating existing precedent, very little new law would be left to future judges to create.¹⁷⁵

In Rasmusen's analysis, the stability of a stare decisis equilibrium requires that judges not overstate their holdings. Otherwise, judges will not afford themselves or their colleagues the opportunity to develop law unencumbered by any prior judge's overreaching in future periods. Overreaching in this analysis occurs when judges claim holding status for assertions that should be dicta, which Rasmusen asserts are not binding "by convention."

Rasmusen's argument suggests both an equilibrium in which rational judges limit their holdings to ensure sufficient room for future law making by other judges and that judges must self consciously restrain themselves so as not to disturb this equilibrium.¹⁷⁶ Thus, using an extreme example, Rasmusen suggests that an ambitious judge might thwart the equilibrium if he were to use a bankruptcy case as a vehicle for a book length treatment on such wide ranging issues as abortion, copyright, and criminal procedure, while presumably claiming holding status for the resulting discussion. The difficulty is that if it is rational for ambitious judges to engage in a strategy of defection, thus producing such excessive opinions, then the game is closer to a prisoners' dilemma in which the dominant strategy would be mutual defection rather than cooperation. While some judges might voluntarily restrain themselves, nothing prevents other judges from assuming a more ambitious law making role, which would cause the claimed equilibrium to unravel.

While the explanations that O'Hara and Rasmusen offer would appear to invite mutual defection as the dominant strategy, refining the models will allow us to identify the basis for a more benign equilibrium outcome. The rational choice analysis will help to explain the emergence of a norm concerning the scope of judicial opinions. This norm encourages judges to limit their holdings based upon the material case facts. Judges are motivated to do so as a means of ensuring their own ability to police against defections by other judges on the same court who would prefer to set out their own policy preferences unconstrained by precedent. A precedent

¹⁷⁵ Rasmusen, *supra* note 144, at 75.

¹⁷⁶ Professor Rasmusen further suggests that the need to avoid inconsistency with external law provides an additional endogenous incentive against overreaching since conflicts in law are likely to rise with ambitious opinions. *See id.* While this is certainly possible, we also think that judges are generally qualified to avoid obvious conflicts with existing law and that, as a result, this concern is unlikely to strongly motivate limiting the scope of judicial opinions.

setting judge will rightly fear that as the precedent setting opinion appears to overreach, that opinion will be increasingly vulnerable to future judges who not only will distinguish the factual predicates that give rise to tangential aspects of the opinion, but also will disingenuously distinguish those parts of the opinion that represent core holdings. In this analysis, the motivation to limit the scope of opinions stems entirely from rational motivation of judges *to enhance* their own power to make law rather than to arbitrate disputes, and since all judges on the same level court share this incentive, it provides the basis for a stable equilibrium outcome, one that in this instance takes the form of a norm, rather than that of a rule or a modification to institutional structure.

Consider a very simple example involving two judges, each operating as single judge appeals court panels, who resolve three cases in sequence. Assume that Judge Potter holds strong views on whether the optimal tort regime is one of contributory or comparative negligence. Further assume that Judge Potter is presented with Case I, a contract dispute that turns solely on whether the claimed contract was supported by consideration. Assume that Judge Potter resolves the contract dispute in favor of the plaintiff, finding consideration, and goes on to state: “I further hold that in a tort suit, the rule of contributory negligence applies such that when plaintiff is more than 50% at fault, defendant is afforded complete relief based upon plaintiff’s comparative fault, but when plaintiff’s fault is 50% or less, defendant is entitled to no offset based upon plaintiff’s comparative fault.”

One year later, Judge Weasley is called upon to resolve Case II, a negligence suit in which the jury, in a special verdict, determined that the plaintiff is 40% at fault. The trial judge, relying upon Case I as a precedent, determined that because the plaintiff was not more than 50% percent at fault, under the Judge Potter’s selected regime of contributory negligence, the defendant was not entitled to any liability offset resulting from plaintiff’s comparative fault. On appeal to Judge Weasley, the defendant claims that Case I was not controlling and that the preferred regime—one Judge Weasley would be free to adopt if Case I is not controlling—is comparative negligence, under which defendant’s liability would be offset by plaintiff’s 40% comparative fault.

In his opinion in Case II, we might anticipate Judge Weasley saying something like the following:

Case I presented a contract dispute and turned on the whether or not consideration was present. This case presents the question whether the governing tort regime is contributory or comparative negligence. While Judge Potter expressed views on the issue presented in this case, and further characterized those views as stating a holding, nothing in Case I turned on his determination on that issue. As a result, regardless of the label that he attached to that discussion, anything expressed on this question in Case I falls squarely into the category of dicta.

In this analysis, Judge Weasley avoids the claimed obligation of precedent by relying upon the material facts in two cases: first, the case in which the claimed holding was articulated, and second, the case presently before him. In addition, by asserting that nothing turned on the tort discussion, Judge Weasley has clarified that the discussion did not lead to Judge Potter's earlier judgment.

Anticipating this type of response by Judge Weasley creates a problem for judges, including Judge Potter, in future cases that they decide. Judge Potter realizes that he cannot affect doctrine outside the context of the material facts that he relies upon to justify his judgment in the case he is actually called upon to decide. But that is not all. He also is concerned that if he is not careful in articulating a defense of his genuine holdings, in the first case involving an issue of consideration in a contract dispute, then future judges might further characterize those holdings as dicta. Unless Judge Potter carefully defends a holding based upon a justification for making law that is grounded in those issues implicated by the material case facts and that explain the basis for his judgment, he runs the risk that Judge Weasley, or other judges, not only will distinguish that which is obviously outside the case, but also disingenuously distinguish core holdings.

The analysis reveals two critical features of the stare decisis norm as it relates to the scope of written opinions. First, it shows that judges can credibly affect only the direction of law in the cases that they are called upon to decide based upon the material facts that lead to the judgment. Second, it reveals that when writing opinions, judges will be rationally motivated to seek to protect their core holdings by grounding those holdings by explaining how the material facts of the case justify the disposition. Otherwise, they will rightly fear that future jurists might attempt not merely to distinguish the parts of their opinions that go beyond the core facts, and which thus constitute legitimate distinctions, but also that overly broad holdings will invite disingenuous distinctions that cut into core holdings. Thus, in a case in which Judge Potter

actually must resolve a series of legal issues to resolve the dispute before him, he will be concerned that any failure to justify the need to reach each issue, by demonstrating a logical progression that links the resolutions of the issues he claims to resolve to the material facts of the case, will invite future jurists to limit the reach of the holdings that are credibly grounded in the fortuitous circumstances of the case.

Conversely, future jurists will read published opinions for material fact-based justifications of any newly made law as a precondition to applying holdings within those opinions as precedent in the cases before them to which no material distinctions apply. Over time, the resulting practice of defending opinions based upon material facts and seeking material fact-based justifications for written opinions provides a stable solution to the potential defection strategy that would otherwise dominate and encourage judges to overreach, if not by writing the proverbial 600 page opinion, at least by extending beyond the material case facts in the resolution of immediate cases. And notice that this equilibrium is not the product of any desire to leave room for future jurists to make law, or of any reticence among judges to make law themselves. Unencumbered by the threatened non-enforcement of core holdings, judges would happily write opinions that go well beyond immediate case facts, or at least those judges who are confident of their own jurisprudential views would. But this model provides a rational incentive to limit defection, namely the desire to ensure that although the law making role is thereby limited to chance presentation of issues linked to material facts, at least any law that jurists make in resolving those facts will be respected as precedent.

Thus far, the analysis reveals that stare decisis forms an endogenous constraint that limits, to some extent, the tendency of appellate judges to overreach in claiming holding status for propositions of law that are not necessary to the resolution of the cases before them. Limiting holdings to those issues the resolution of which can be credibly defended as necessary to resolve a case based upon the material facts articulated in a judicial opinion allows appellate judges to increase the impact of their decisions. More broadly, voluntary self-imposed limits on the scope of holdings allow the appeals courts to move from institutions that resolve disputes to institutions that make law in the course of performing that task. But the very transformation from arbitrators to jurists invites a tendency to overreach precisely because judges, unlike arbitrators, have substantial power to make law.

Other factors, including concerns for reputation and collegiality, also influence decisions to cast holdings broadly or narrowly, and certainly we do not suggest that our simple rational choice model captures all relevant concerns. For example, once we recognize that appeals courts do not operate in single judge panels, and that majorities on panels of three for most appeals and of full appellate courts for en banc reconsiderations are necessary to create precedent, we can further appreciate that strategy might sometimes overtake candor as the basis for decisions to join with others as a means of creating desired precedent.¹⁷⁷ While rational choice helps to answer some important questions concerning why judges do not systematically overreach, the analysis thus far is at best a rough cut, one that provides a useful starting point in undertaking a careful normative analysis of dicta and holding. This is especially so once we realize that most cases involve facts or law that are neither obviously within nor beyond the scope of what is needed to resolve a case. Having identified some core limits exposed by the rational choice model, we now turn to four overlapping sets of normative concerns, which prove critical in defining dicta and holding.

B. The Judicial Four C's

In this section, we identify four normative considerations, which we will call the “Judicial Four C’s.” Jewelers evaluate carat, cut, color, and clarity in assessing diamonds; we evaluate constraint, consideration, clarity, and candor in forming understandings of what to treat as holding and dicta within judicial opinions. While our considerations only overlap directly in one category, clarity, this area of overlap proves particularly significant. As with diamonds, the principal difficulty in overly broad holdings is that of inclusions, and specifically of clouding what should be clear statements of the holding with other assertions that should instead have been treated as dicta. While we do not employ cut as a criterion, just as a diamond’s cut can be too shallow or too deep, the factors that we employ do not push inexorably in a single direction.¹⁷⁸ In the case of judicial consideration, for example, this means that judges resolve those issues targeted to the immediate case, without avoiding difficult questions necessary to resolving cases presented, but not inquiries that cases might implicate if conceived at a broader

¹⁷⁷ For discussions of strategic versus candid behavior on multi-member appellate courts, see STEARNS, *supra* note x, at 97-156 (2002); Evan Caminker, *Sincere and Strategic Voting Norms on Multimember Courts*, 97 MICH. L. REV. (1999); and Maxwell L. Stearns, *Should Justices Ever Switch Votes?: Miller v. Albright in Social Choice Perspective*, 7 S. CT. ECON. REV. 87 (1999).

¹⁷⁸ For information concerning diamond grading, see http://www.gia.edu/about/33/what_is_gia.cfm.

level. The criteria are useful, of course, not only because they might allow us to grade judicial opinions, but because they allow us to assess possible definitions of dicta, and more specifically the analytical categories and problems that we identified in Part II.

We recognize that the normative considerations we consider not only push in more than a single direction, but also that they are incommensurate. Weighing these factors against one another requires us to make normative judgments that some might dispute. In addition, although we doubt that many readers would claim that any of these four factors are unimportant, we recognize that some might place different weights on these factors, and might include other normative factors as well. We believe that each of the considerations we have chosen is relatively uncontroversial in its own right.

1. Constraint

Constraint is perhaps the most significant normative consideration in justifying limits on the scope of holdings. At first blush, this concern might appear to push solely in a single direction. Limiting holdings to the resolution of issues that are implicated by the material case facts substantially constrains the power of judges to make law on extraneous issues. Assuming agreement on what constitutes extraneous discussion, such discussion should be understood as dicta. But even this is not so simple. The doctrines of ripeness, standing, and mootness, though frequently deployed in the name of judicial constraint, are often criticized as devices that allow an unrestrained judiciary to avoid deciding, or to choose when to decide, some of the most difficult questions of law and public policy that are implicated in properly presented cases.¹⁷⁹ Constraint therefore is relevant in assessing not only when a court has strayed too far, but also when it has decided to avoid going down a necessary, but difficult or unpleasant, path.

Along with the remaining considerations, constraint is best understood at a micro level, one that focuses not on the role of the case against some broad set of doctrines, justiciability or otherwise, but rather on the scope of judicial autonomy in deciding the particular case. Even if the rules of the judging game impose considerable restraint, within individual cases, judges

¹⁷⁹ For a neofederalist critique of standing that is consistent with this analysis, see Robert J. Pushaw, *Justiciability and Separation of Powers: A Neo-Federalist Approach*, 81 CORNELL L. REV. 393 (1996). For a social choice analysis responding to various criticism of standing, including the claim that it is primarily an avoidance device, see Maxwell L. Stearns, *Standing Back from the Forest: Justiciability and Social Choice*, 83 CAL. L. REV. 1309 (1995), which develops a theoretical model, and Maxwell L. Stearns, *Standing and Social Choice: Historical Evidence*, 144 U. PA. L. REV. 309 (1995), which presents historical evidence in support of the model.

retain substantial flexibility. And indeed, as we will argue in the next part, we believe that the definitions of dicta and holding must be sufficiently capacious to allow the deciding jurists such flexibility in setting out the reasoning within particular cases. Consistent with our earlier positive analysis,¹⁸⁰ however, we also believe that the legal system should restrain judges from conclusively resolving issues not meaningfully presented by the material facts of the particular case.

Some might argue that courts have no less a role in setting out legal policy than legislatures, and therefore that there are no inherent limits linked to the obligation to resolve particular cases on judicial law making powers.¹⁸¹ We respectfully disagree. Courts inevitably make positive law, but that does not mean that individual judges are normatively justified in making positive law at will. A central feature of the judicial system is the random selection of judges to hear cases and thus to resolve the issues presented by those cases.¹⁸² Cases arise from circumstances beyond the control of the individual judges, thus limiting any given judge's ability to determine which issues to resolve.¹⁸³ Proper definitions of holding and dicta can preserve this feature while still allowing judges control over case reasoning on issues fairly presented. To the extent that judges may resolve issues independent of the random generation of cases in need of resolution and of the factual bases of the cases that arise, a central aspect of judicial legitimacy—and thus of judicial restraint—is undermined.

Although legal scholars might disagree as to how much constraint is desirable, even those who believe that judges should have a substantially stronger policymaking role recognize the need for some restraint on judicial lawmaking.¹⁸⁴ Imagine a judicial system in which there were no constraint, whether taking the form of formal rules or informal norms, encouraging judges to

¹⁸⁰ See *supra* Part III.A.

¹⁸¹ For an argument from the left that would relax modern conceptions of justiciability as a precondition to judicial lawmaking, see Mark V. Tushnet, *The New Law of Standing: A Plea for Abandonment*, 62 CORNELL L. REV. 663 (1997), and for an argument from the right that would relax modern standing concepts to allow a coequal role of the federal judiciary with Congress, see Pushaw, *supra* note x.

¹⁸² See Michael Solimine, *Nepotism in the Federal Judiciary*, 71 U. CINN. L. REV. 563, 576 (2004) (describing random assignment as furthering notion of intrajudicial equality). Even in one-judge jurisdictions, the judge is constrained to decide only the cases that litigants bring.

¹⁸³ See STEARNS, *supra* note x, at 201-02 (positing that judicial lawmaking is legitimated by the ad hoc and as needed circumstances of resolving issues in the context to deciding cases).

¹⁸⁴ See, e.g., William N. Eskridge, Jr., *Politics Without Romance: Implications of Public Choice Theory for Statutory Interpretation*, 74 VA. L. REV. 275, 307 (1988) (cautioning against “free inquiry approach” to statutory interpretation while advocating a generally dynamic approach relative to actual practice); Pushaw, *supra* note 179, at 399 (arguing for a relaxation of justiciability doctrines to facilitate greater judicial check on elected branches within general framework constrained by principles of separation of powers).

limit their case dispositions to resolving issues implicated by material case facts. Any judge then would be free to resolve—and indeed might feel justified in resolving—any issue at any time the he or she chose. This judicial system then would be akin to a legislature in which any single legislator could pass a bill.¹⁸⁵ Other rules or norms still might constrain judges relative to legislators—for example, *stare decisis*, the custom of honoring legislative intent, and the expectation that judges explain their reasoning in written opinions—but such a system would certainly not further, and might undermine, these practices. The judges who would elect to resolve particular issues would likely be those who care most deeply about the issues they elect to consider. In electing to resolve issues not obviously implicated in the cases before them, they might willingly sacrifice constraint to ensure preferred outcomes.

We recognize that there are some potential virtues in affording judges broad flexibility in identifying on their own those issues that they elect to resolve. Granting broad flexibility might promote valuable specialization. The judge who cares most about admiralty law is likely to be the judge who knows the most about admiralty law, and such a judge might be particularly eager to draft a broad rather than a narrow holding in an admiralty case, one that anticipates a problem important to that body of precedent. With the specialist, rather than a dabbler, deciding such an issue, there might be a lower risk that the failure to appreciate the peculiarities of admiralty law will generate problematic doctrine. Similarly, more capable judges might be better able to identify remaining legal ambiguities in doctrine than less capable judges, and a system that effectively allows judges to resolve cases or issues within cases more broadly depending upon their abilities might generate sounder legal doctrine.

Again, however, we must consider important countervailing concerns. By allowing the specialist—whether in admiralty, bankruptcy, corporate law, tax law, or something else entirely—to drive precedent in a chosen area, the judicial system would create discrete doctrinal enclaves that might not reflect broader doctrinal or institutional concerns. We might then confront a choice between a system with numerous bodies of specialty judges and specialty jurisprudence, versus a system of generalists who encourage various specialized bodies of law to conform more comfortably with an overriding set of agreed upon jurisprudential principles.¹⁸⁶

¹⁸⁵ For an analysis that relies upon social choice to draw these distinctions between judicial and legislative decision making, see Maxwell L. Stearns, *The Misguided Renaissance of Social Choice*, 103 *YALE L.J.* 1219, 1288-89 (1994).

¹⁸⁶ For an essay defending the generalist appellate judge, see Richard A. Posner, *Will the Federal Courts of Appeals Survive Until*

Moreover, a legal system that allowed judges unfettered discretion to determine what issues to resolve might invite ideology as a determinant of which issues particular judges elected to address. Legal doctrine might become less coherent with ideologically polarized judges simultaneously resolving issues in a particular substantive area.¹⁸⁷ We do not suggest that specialization should be altogether avoided, but rather that there are substantial tradeoffs in a system that invites systematic specialization among judges who might care more deeply about particular outcomes or bodies of case law than with reconciling that case law within the larger context of the legal system.

The challenge is for the legal system to discourage judges from reaching out to resolve issues, while allowing them to resolve, and perhaps even demanding that they resolve, those issues that do arise. The legal system might grant a sufficient amount of flexibility in identifying issues so as to allow more specialized and more intelligent judges a greater role than other judges, without allowing undue influence of ideology on who decides any given issue. In economic terms, the legal system should allow flexibility up to the point at which the marginal costs of flexibility exceed the marginal benefits. We will argue that the proper distinction between holding and dicta affords judges considerable latitude to select among various paths that lead from facts to judgment, and among levels of breadth, in forging a particular path of case reasoning.¹⁸⁸ Our definitions, however, will also prevent judges from reaching out to resolve issues on paths that do not lead from facts to judgment. A judge can retain considerable flexibility even while restrained by the rules and norms of the judicial game.

2. *Consideration*

In a legal system in which a judge could use any case as the vehicle through which to resolve issues of her choosing, some judges might race to resolve particular issues for which they hold particularly strong normative concern. Not only would such judges be less likely to be neutral, but also the race itself might reduce the care with which they address issues. The American legal system includes various mechanisms that encourage proper consideration before judges resolve those issues implicated in the cases they are deciding. These include finding

1984? *An Essay on Delegation and Specialization of the Judicial Function*, 56 S. CAL. L. REV. 761 (1983).

¹⁸⁷ See, e.g., Posner, *supra* note 186, at 783 (giving bankruptcy and criminal law as areas in which judges tend to be ideologically polarized).

¹⁸⁸ See *infra* Part IV.B (defining dicta and holding).

issues not previously litigated to be waived on appeal,¹⁸⁹ discouraging (although not prohibiting) *ex parte* adjudication,¹⁹⁰ and the various justiciability rules—ripeness, standing, and mootness—that limit access to the judicial process absent the requisite circumstances to ensure that consideration takes place within a concrete judicial setting.¹⁹¹ At the same time, the expectation that judges explain how they reached their decisions encourages judges to pay considerable attention to those issues that are properly before them. The holding-dicta distinction similarly should serve both functions, delaying resolution of issues until judges can properly consider them and encouraging judges to focus on the issues before them.¹⁹²

The propositions that judges should be hesitant to create new law absent adequate presentation, normally through the adversarial process, and that judges are required to exercise proper care in resolving the cases they decide, are, of course, not new. To the contrary, they are emblematic features of judicial law making.¹⁹³ Cass Sunstein, for example, has recognized the value of judicial minimalism and what he calls “incompletely theorized agreements.”¹⁹⁴ Judges should seek agreement on how to resolve a case, even if that agreement masks chasms on larger

¹⁸⁹ See, e.g., *Carducci v. Regan*, 714 F.2d 171, 177 (D.C. Cir. 1983 (“[W]here counsel has made no attempt to address the issue, we will not remedy the defect, especially where, as here, important questions of far-reaching significance are involved.”) (internal quotation marks omitted).

¹⁹⁰ Of course *ex parte* cases can and do produce important precedents. For perhaps the most famous example, see *Marbury v. Madison*, 5 U.S. (1 Cranch) 137 (1803).

¹⁹¹ Indeed, in *Baker v. Carr*, 369 U.S. 186 (1962), the Supreme Court identified the goal of ensuring the crisp presentation of issues as a central justification for standing, stating:

Have the appellants alleged such a personal stake in the outcome of the controversy as to assure that concrete adverseness which sharpens the presentation of issues upon which the court so largely depends for illumination of difficult constitutional questions? This is the gist of the question of standing.

Id. at 204.

¹⁹² Judge Posner focuses on the latter of these concerns in an opinion in which he addresses the definition of dicta:

[I]nstead of asking what the word “dictum” we can ask what reasons there are against a court’s giving weight to a passage found in a previous opinion. There are many. One is that the passage was unnecessary to the outcome of the earlier case and therefore perhaps not as fully considered as it would have been if it were essential to the outcome. A closely related reason is that the passage was not an integral part of the earlier opinion—it can be sloughed off without damaging the analytical structure of the opinion, and so it was a redundant part of that opinion and, again, may not have been fully considered. Still another reason is that the passage was not grounded in the facts of the case and the judges may therefore have lacked an adequate experiential basis for it; another, that the issue addressed in the passage was not presented as an issue, hence was not refined by the fires of adversary presentation. All these are reasons for thinking that a particular passage was not a fully measured judicial pronouncement, that it was not likely to be relied on by readers, and indeed that it may not have been part of the decision that resolved the case or controversy on which the court’s jurisdiction depended (if a federal court).

United States v. Crawley, 837 F.2d 291, 292-93 (7th Cir. 1988). We agree with Judge Posner’s analysis, but we do not believe that consideration is the only relevant factor.

¹⁹³ The virtue of ensuring adequate consideration, however, has received renewed attention in recent years, particularly in CASS R. SUNSTEIN, *ONE CASE AT A TIME: JUDICIAL MINIMALISM ON THE SUPREME COURT* (1999). Sunstein defines the practice of “decisional minimalism” as “the phenomenon of saying no more than necessary to justify an outcome, and leaving as much as possible undecided.” *Id.* at 3.

¹⁹⁴ Cass R. Sunstein, *Incompletely Theorized Agreements*, 108 HARV. L. REV. 1733 (1995).

jurisprudential or philosophical issues that the case may present.¹⁹⁵ The holding-dicta distinction, properly conceived, helps to enforce an appropriate level of consideration, one that delves into reconciling the case with the larger body of precedent, but that does not necessarily look for, or attempt to resolve, fissures within larger bodies of case law, at least at one time. Judicial minimalism leaves important issues open for resolution after further democratic deliberation.¹⁹⁶

3. Clarity

Once a judge has fully considered a particular issue, the judge must still decide whether to offer a broad or narrow resolution of the issue. Judicial minimalism often calls for narrow resolutions. In some cases, however, “a wide rule, even if overinclusive and underinclusive, would be better than a narrow judgment,” Sunstein acknowledges, because in the absence of a wide rule “lower courts and subsequent cases would generate an even higher rate of error.”¹⁹⁷ A legal regime that insisted that issues always be resolved as narrowly as possible might invite paralysis. There is value in achieving legal clarity, and although clarity sometimes pushes in the same direction as do the factors of constraint and consideration, other times clarity pushes in the opposite direction.

Broad rulings can promote clarity by cutting off a subsequent line of cases as, for example, would have occurred had the Supreme Court ruled for the state in *Roe v. Wade*,¹⁹⁸ and found no constitutional right to abort.¹⁹⁹ Courts sometimes promote doctrinal clarity, however, by engaging in the opposite strategy, thus providing very specific and detailed holdings. Justice Powell’s opinion in *Bakke*, for example, gave educational institutions a relatively clear

¹⁹⁵ Thus, Sunstein states:

[W]ell-functioning legal systems often tend to adopt a special strategy for producing agreement amidst pluralism. Participants in legal controversies try to produce incompletely theorized agreements on particular outcomes. They agree on the result and on relatively narrow or low-level explanations for it. They need not agree on fundamental principle. They do not offer larger or more abstract explanations than are necessary to decide the case. When they disagree on an abstraction, they move to a level of greater particularity. The distinctive feature of this account is that it emphasizes agreement on (relative) particulars rather than on (relative) abstractions. This is an important source of social stability and an important way for diverse people to demonstrate mutual respect, in law especially but also in liberal democracy as a whole.

Id. at 1735-36.

¹⁹⁶ See SUNSTEIN, *supra* note 193, at 24-25. For an argument countering that minimalism can produce avoidance, see Pushaw, *supra* note x.

¹⁹⁷ SUNSTEIN, *supra* note 193, at 49-50.

¹⁹⁸ 410 U.S. 113 (1973); *infra* Part V.B.

¹⁹⁹ For a list of 46 post-*Roe* cases involving abortion, counting only up to 1993, see Paul Benjamin Linton, Planned Parenthood v. Casey: *The Flight from Reason in the Supreme Court*, 13 ST. LOUIS U. PUB. L. REV. 15, 77 n.283 (1993).

framework for developing affirmative action policies, or at least appeared to have done so from the time of the initial decision in 1978 until at least 1995,²⁰⁰ by signaling his approval of the Harvard plus-factor plan, even though that plan was not immediately before the Court.²⁰¹

4. *Candor*

The clearer the holding-dicta distinction itself, the more difficult it will be for judges to offer disingenuous rationales for avoiding the central determinations of previous cases. As the distinction between holding and dicta becomes increasingly vague, past precedents can be increasingly manipulated. Judges will face greater temptation to cheat from the regime of mutual cooperation,²⁰² or perhaps even convince themselves that they are not cheating, when they can offer some facially plausible argument for disregarding a statement in a prior case. If the holding-dicta distinction were perfectly clear (a goal that we recognize as impossible), then disingenuous manipulation of precedents would be immediately recognizable. That would reduce the incidence of manipulation and improve the legitimacy of the judicial process. Our goal is to move the judicial system closer to this ideal, thus improving judicial candor.

By candor, we do not mean a naïve sense of disclosing the jurist's subjective sensibilities about legal doctrine or case facts. Rather, we mean that the legal system seeks to ensure that judges disclose their genuine views concerning how the case they are deciding should be resolved within the context of existing legal doctrines. It is entirely possible that if the deciding jurists had the power to define the governing legal doctrines without regard to precedent, they would do so differently. Perhaps the most significant benefit of candor in the sense that we use that term is in furthering reliance upon the rule of law. Absent meaningful expositions on the manner in which the court applies the governing rules to the case facts in the course of reaching a

²⁰⁰ See *infra* notes 312-318 (discussing *Hopwood v. Texas*, 732 F.2d 932 (5th Cir. 1996)).

²⁰¹ Perhaps ironically, Lino Graglia, an outspoken opponents of affirmative action residing at the University of Texas Law School, recently supported this point:

A task this difficult [namely reconciling affirmative action in higher education with the principal of nondiscrimination] requires superior intellect, such as one could expect to find at Harvard, and Harvard, Powell thought, had in fact already pulled it off. All that remained to be done was raise Harvard's achievement to the status of constitutional law, which is exactly what Powell's decisive opinion in *Bakke* did. . . . We are being indirectly ruled by the Harvard faculty, however, *Bakke* illustrates with unusual clarity through the ruse of the Supreme Court—mirror and mouthpiece of liberal elite academia—purporting to interpret the Constitution.

Lino Graglia, *Grutter and Gratz: Race Preference to Increase Racial Representation Held "Patently Unconstitutional" Unless Done Subtly Enough in the Name of Pursuing "Diversity,"* 98 TUL. L. REV. 2037, 2040 (2004).

²⁰² See *supra* Part III.A.

disposition, it would be difficult for judges, or other affected actors, to discern with a reasonable degree of certainty the rule of law as it applies in the next related dispute.²⁰³

Judicial incentives for candor, of course, depend not only on the clarity of the holding-dicta distinction, but also on how broadly holding is defined. Again, however, there are cross-cutting concerns. On one hand, a narrower definition of holding, by granting future judges more flexibility, may also encourage judges to be more candid. The power to create precedents, we have seen, may encourage judges to interpret precedent other than in a purely neutral manner. It is, after all, a common pastime for legal academics to point out exercises in creative judicial draftsmanship,²⁰⁴ sometimes with admiration (depending on the importance of the doctrinal objective), but often with a smirk. A definition of the holding-dicta distinction that allows great flexibility for judges to craft broad holdings may also increase the incentives that later judges have to offer disingenuous reasons for avoiding holdings. Even if it is theoretically desirable to have a regime allowing broad holdings, such a regime might bring about its own collapse.

On the other hand, a narrow definition of holding may make judges less willing to express nuanced legal conclusions. If only part of a judge's reasoning will constitute the holding, a judge might try to articulate a holding in a broader and less nuanced fashion than the judge would ideally prefer. More worrisome still, to the extent that the judge believes the nuances critical to the analysis, the judge might adopt a position opposite the judge's preferred case resolution.²⁰⁵ By switching positions and sacrificing the judge's preferred resolution in the case at hand, the judge may be able to ensure that caveats that otherwise would be considered dicta will count as holding, although at the expense of too broad a holding in the opposite direction.²⁰⁶ We believe that a regime that encourages a judge to disguise true beliefs about cases ultimately

²⁰³ For a related analysis demonstrating that vote trading within appellate courts would undermine reliance on the rule of law, see STEARNS, *supra* note x, at 88-92.

²⁰⁴ See, e.g., Dorf, *supra* note x, at 2021-24 (criticizing the Supreme Court's recasting of earlier executive removal precedents in *Morrison v. Olson*, 487 U.S. 654 (1988)).

²⁰⁵ Some doctrines, such as the Supreme Court's outcome voting regime, discourage judges from endorsing case resolutions with which they disagree as a means of achieving some other goal. See Michael Abramowicz & Maxwell L. Stearns, *Beyond Counting Votes: The Political Economy of Bush v. Gore*, 54 VAND. L. REV. 1849, 1932 (2001) (demonstrating that social choice "reveals one important constraining effect [the outcome voting] rule places on judicial behavior by encouraging not only principled—meaning nonstrategic—articulation of issues, but also principled resolution of issues").

²⁰⁶ For a related analysis using social choice to demonstrate that the narrowest grounds rule limits strategy in *Bakke* by preventing Justice Powell from being forced to select an opinion that strikes down the Davis plan but that employs an overly stringent understanding of the permissible use of race in admissions (per Justice Stevens), or an opinion that upholds the Davis plan, thus permitting the use of race in admissions, but in a manner that Powell found constitutionally defective (per Justice Brennan), see STEARNS, *supra* note 38, at 137-38.

undermines the rule of law, first by reducing predictability and legal clarity, and second by inhibiting the emergence of nuanced doctrine.

C. *Assessing Propositions and Problem Categories*

We will now apply the Four C's to the various categories of potential holdings that we defined in Part II.

1. *Necessary Propositions*

The most straightforward category to analyze is necessary propositions, such as the proposition in *Bakke* that the Equal Protection Clause protects nonminorities.²⁰⁷ Unless a system of adjudication rejects altogether the notion that legal principles can constitute holdings,²⁰⁸ a proposition that is necessary to the disposition (or, by extension, to another holding) itself must count as a holding. Absent an overruling, a subsequent court must distinguish a precedent if it is to arrive at an opposite result. And unless the judicial system is willing to invite upon itself claims of complete disingenuousness, a subsequent court will need to reconcile its ruling with the earlier case.²⁰⁹ Denying relief to the next white applicant on otherwise indistinguishable facts because Cotter begins with the letter “C” while *Bakke* begins with the letter “B” will not do. If we set aside such obvious absurdities, it is apparent that no credible court could reconcile a holding resting upon the determination that a racial set aside is not a race-based classification or that Equal Protection Clause does not protect whites with the ultimate holding in *Bakke*. Under any fair reading of the case, *Bakke* could not have prevailed without succeeding on these two arguments.²¹⁰ Where a Court explicitly endorses a proposition and could not have reached the

²⁰⁷ It is, of course, our ambition to assess whether propositions are holdings without regard to whether the propositions are correct or persuasive. We recognize that some might conclude that the Equal Protection Clause properly understood does not protect whites. See *The Slaughter-House Cases*, 83 U.S. (16 Wall.) 36 (1873) (construing the Equal Protection Clause as primarily motivated by the need to protect blacks against state action). But if precedent has any constraining influence at all, a court determining whether to apply precedent must consider more than simply whether it would have resolved the initial case differently. See *Planned Parenthood v. Casey*, 505 U.S. 833 (1992) (listing factors to consider in applying stare decisis and applying those factors to avoid overruling *Roe v. Wade*, 410 U.S. 113 (1973)).

²⁰⁸ For discussion of legal systems without precedent, see *supra* note 122; and *infra* note 265.

²⁰⁹ In *Bakke*, Statement (7) (the Davis program, which focuses solely on race, hinders rather than furthers diversity) is the narrowest possible expression of the holding, applying itself only to the immediate case. An interesting question is what the effect of (7) would be if the Court had simply announced the facts and then skipped immediately to (7) without explaining its justification. We would argue that in such a case, other courts would still have to justify any opposite results from the result announced by (7), by identifying a relevant distinction in the facts. See *infra* Part IV.A.1 (arguing that the even in the absence of any articulated reasoning, the combination of facts and the announced judgment create a very narrow holding).

²¹⁰ As this analysis suggests, the necessity of considering and resolving an issue as a court resolved it is sufficient for the court's resolution of the issue to count as a holding. But necessity is not required for a statement to be a holding. See *infra* Part IV.A.3. In

result without at least implicitly endorsing the proposition, then the proposition must be a holding.

Although we recognize that this will be the least controversial of our conclusions, it is worth establishing that the four normative factors that we have identified strongly support the categorization of necessary propositions as holdings, at least absent other concerns. Where a court *must* resolve an issue to resolve a case, and in fact does resolve that issue, the court cannot be thought of as having reached out to resolve the issue, so concerns about constraint do not arise. The failure to resolve the issue, although potentially ensuring continued consideration of the issue over time, would entail a failure to dispose of the immediate case based upon a proper consideration of a critical issue. Perhaps most important, counting such propositions as holdings is essential to maintaining legal clarity, for the set of holdings would be very small if necessary propositions did not count. Finally, concerns about candor seem irrelevant. The resolution of a necessary proposition does not alter the need for a judge to resolve any other propositions, and so judges can be expected to resolve these questions based on their candid views rather than based on strategic considerations.

2. Sufficient Propositions and Related Problems

We also conclude that sufficient propositions for the disposition (or, by extension, for other holdings) should at least presumptively be holdings, and that the existence of alternative possible justifications or alternative justifications should not be a bar to crediting sufficient propositions as holdings. Let us consider alternative possible justifications first. The principal concern is that when alternative possible justifications exist, we know that the jurist has made some choice. It is possible that a judge might use such flexibility as a vehicle through which to make law, and judges with strong preexisting normative views on particular issues may be more likely to resolve those issues than other judges. To that extent, alternative possible justifications might trigger concerns about constraint.

While we believe that the normative merit of creating law within cases is a function of the judge's need to resolve cases, we do not believe that judges should be constrained in choosing among decisional paths when the legal doctrines implicated in the cases they decide

addition, the logical necessity of deciding an issue that the court has not actually decided might not transform a potential holding into a holding. *See infra* Part IV.B.1.

require the exercise of judicial discretion. One might criticize particular legal doctrines for affording judges too much leeway, but some leeway is inherent in the judicial process. If a test requires that two prongs be met to achieve a result, a judge cannot be said to overreach in finding that because one of the two elements is not met, the test is not met. The judge must pick some decisional path to resolve the actual case presented. As long as a judge does pick an appropriate path from material facts to judgment, the judge remains considerably constrained.

Two other factors, meanwhile, are decisive in suggesting that alternative possible justifications should not bar sufficient propositions from counting as holdings. First, if alternative possible justifications exempted actual justifications from holding status, it would be difficult to develop clarity in the law. There almost always exist alternative possible justifications for the result that the court reached with respect to a dispositive issue or to the case as a whole.²¹¹ Second, a holding rule excluding alternative possible justifications would lead judges determined to establish a holding to consider the full range of alternative possible justifications. Because there may be a large number of alternative possible justifications, this pursuit would call the judge to consider a large number of issues, distracting the judge from focused consideration of any one of those issues.

The analysis of cases in which a judge in fact offers alternative justifications for the same result is equally straightforward, indeed easy given the conclusion that alternative possible justifications should not deprive sufficient propositions that the court did reach of presumptive holding status. There exists some risk that a judge will offer two justifications rather than one because the judge is particularly concerned about making law with respect to both issues, thus once again exposing concerns about constraint. The concern, however, can only be with respect to the issue about which the judge cares *less*. Give our resolution of the alternative possible justification problem, a judge would have the flexibility to resolve at least one of two alternative justifications regardless of the rule governing alternative justifications. Granting the judge the power to resolve a second (or third or fourth) alternative justification does not seem particularly

²¹¹ As we noted in *Bakke*, for example, the Court conceivably might have struck down the program on the basis of sex discrimination rather than race discrimination. *See supra* notes x-x and accompanying text. That this seems exceedingly unlikely under the facts of *Bakke* may suggest that the real question is whether there were alternative *plausible* justifications. That formulation, however, emphasizes the difficulty of assessing the reasonableness of justifications that have not been explored. The holding-dicta distinction would not be clear if it demanded an analysis of legal issues that the court did not even consider. Alternative possible justifications might be found only where litigants in fact advanced those justifications, but with such a rule, there might be considerable debate about whether litigants did sufficiently raise justifications, an issue that would require looking beyond the judicial opinions themselves.

likely to channel resolution of issues to the judges who feel particularly strongly about those issues.

Denying holding status to alternative justifications would delay the resolution of legal issues for reasons having nothing to do with the need for more consideration, thus rendering the law less clear without a compelling justification. This would thus undermine both the goal of encouraging judges to consider issues genuinely presented in cases, and the goal of promoting clarity in the law. The concern for candor, meanwhile, furnishes an additional reason for counting alternative justifications as holdings. In a regime that denied the status of holding to alternative justifications, a judge might have an incentive to resolve more than one potential justification as a means of *denying* holding status to the proposition establishing the most critical justification in the case. Suppose that Powell had wanted to strike down the Davis program, but did not want to create any precedent on the strict scrutiny analysis of social remediation.²¹² Powell might explain that each prong was sufficient because he wanted neither to be a holding. Because judges establish precedent in resolving cases, the risk that judges will decide cases based primarily on their idiosyncratic predilections toward particular litigants is reduced.

Candor concerns would be particularly severe if both alternative possible justifications and alternative justifications were excluded from status as holding. Imagine a holding rule demanding that whenever an alternative possible justification can be identified, the chosen justification counts as a holding only if the alternative possible justification is expressly considered and rejected. A judge then might have an incentive to reject rather than accept the alternative possible justification simply as a means of ensuring that the actual justification would have the status as holding. We posit that this might cause only slight damage to precedent, because the rejection of the alternative possible justification itself would be dicta under the analysis that follows. A judicial system that invited such manipulation, however, not only would undermine the legitimacy of the judicial process, but also would aggravate the concerns about

²¹² It may seem implausible that Powell would want to strike down the Davis program for reasons other than those actually considered in the case. But there are at least some cases in which judges might want a particular litigant to win, perhaps because of the underlying case equities, but not want to create the precedent that would be necessary for that litigant to win. The controlling opinion in *Bush v. Gore*, 531 U.S. 98 (2000), suggested that it might have viewed that case in this manner, stating: “Our consideration is limited to the present circumstances, for the problem of equal protection in election processes generally presents many complexities.” *Id.* at 109.

lack of constraint, as the judges most eager to set a precedent would be most willing to jump through the necessary hoops to do so.²¹³

3. *Nonsupportive Propositions*

In contrast, nonsupportive propositions generally should count as presumptive dicta. Where a judge resolves an issue in a manner that does not contribute to the disposition of the case, there is a strong possibility that the judge is reaching out to resolve that issue because the judge holds relatively strong views about it. It is one thing to allow a judge to choose among alternative paths of decision leading to the case judgment and to credit issues that are necessary or sufficient to supportive propositions along that path. It is another to allow a judge to choose decisional paths that do *not* lead to the disposition and to credit as holdings the resolution of issues that lead to a place that the court is not compelled to go.

Not all nonsupportive propositions are equally irrelevant. We have distinguished, for example, between absolutely irrelevant propositions (those that could have no conceivable connection to the case) and six categories of nonsupportive propositions that might have been relevant if they had been resolved differently or if other propositions had been resolved differently.²¹⁴ For example, the issue in *Bakke* concerning whether diversity is a compelling interest was at least potentially relevant in the sense that a negative conclusion would have been sufficient for the disposition, and an affirmative conclusion could have been supportive along a path to an opposite case disposition. Thus, although the issue was not on the decisional path to the actual disposition of the case, it was on a decisional path to a potential resolution.

Nonetheless, we believe that the constraint factor is ordinarily sufficient to render resolution of such an issue dicta. While it is certainly not uncommon for a judge to indicate that certain issues point in the direction opposite the disposition of a case, a judge can just as easily skip such issues altogether, and often judges do skip such issues. A judge might use the process

²¹³ The recognition that a regime that fails to credit actual justifications because of alternative possible justifications would produce incentives for noncandid resolution of issues also shows that such a regime would encourage judges to consider issues unnecessarily. Again, using *Bakke*, consider the implications of not crediting the alternative explored justification on societal remediation. Such a regime would insist that Powell specifically rule on both prongs to have even the chance of holding, even though only one of the two prongs is needed for him to reject this justification for the use of race. Justice Powell then would be reaching at least one issue that is inessential to his ultimate disposition. Requiring a judge to resolve an issue the resolution of which is beside the point certainly promotes consideration of that issue, but it does so in the precisely the manner that excursions into dicta promote consideration of matters extraneous to the resolution of the case at hand.

²¹⁴ See *supra* Part II.A.2.b.i.

of writing an opinion to determine how to resolve a case, but more often a judge considers all of the issues in the case before deciding at least tentatively how to structure the opinion.²¹⁵ As recent cognitive psychology research has revealed, judges are like other decision makers in this respect; human reasoning tends not to be linear, with premises leading to conclusions, but bidirectional, with anticipated or preferred conclusions also affecting premises.²¹⁶ Thus, a judge who reaches a nonsupportive proposition typically recognizes that the proposition is nonsupportive before the judge writes the relevant section of the opinion. When a judge decides to resolve that proposition nonetheless, there is a strong possibility that the judge cares especially about the resolution of that issue, and the constraint factor presents a strong argument for counting the resolution of such a proposition as dicta.

The issue remains close, however, because the remaining normative concerns move largely in the other direction. The parties in a case have had incentives to brief potentially relevant issues. Karl Llewellyn recognized long ago that the consideration factor does not necessarily militate in favor of delaying resolution of potentially relevant issues.²¹⁷ On the other hand, a holding-dicta distinction that encourages judges to travel down only decisional paths that actually lead to the disposition will likely encourage greater consideration of issues along that path, rather than diverting judicial attention to issues that arise along several possible alternative paths, some of which might not lead to the judgment. Clarity in the law sometimes might be advanced by resolving issues that have been briefed, but allowing potentially relevant propositions to count as holdings might require difficult distinctions about which propositions in

²¹⁵ Judges sometimes speak of situations in which an opinion “won’t write,” meaning that the process of writing causes the judges to change their minds. See Dan Simon, *A Psychological Model of Judicial Decision Making*, 30 RUTGERS L.J. 1, 115 (1998) (asserting that “[s]ome judges report that on occasion, they find a decision that has wandered too far, or the decision simply “won’t write”). That may happen when the judge realizes that the opinion “does not follow accepted rules and is therefore arbitrary in result or superficial in reasoning.” *Id.*

²¹⁶ See Dan Simon, *A Third View of the Black Box: Cognitive Coherence in Legal Decision Making*, 71 U. CHI. L. REV. 511, 529-83 (2004) (reviewing the cognitive psychology literature and offering legal implications).

²¹⁷ Karl Llewellyn poses the following hypothetical: “Again, take a case where the court rules on four points in favor of a man who won below, but reverses, for all that, on the fifth point.” KARL N. LLEWELLYN, BRAMBLE BUSH: ON OUR LAW AND ITS STUDY 46 (1929). Llewellyn recognizes that the issue is close:

One of the reasons, of the sound ones, often given for weighing dicta lightly, is that the background and consequences of the statement have not been illumined by the argument of counsel, have not received, as being matters to be weighed with brows-a-wrinkle, the full consideration of the court. In the case put the first reason does not fit; the second, if it is to be put on at all, hangs loose and flaps.

Id. Llewellyn’s subsequent analysis, however, focuses not on the fact that the four are nonsupportive propositions, but on the fact that there are five propositions in all. See *id.* at 46-47. Llewellyn thus failed in this portion of his argument to distinguish the problem resulting from nonsupportive propositions and the problem of alternative justifications. Shortly before, however, Llewellyn identifies the category of nonsupportive propositions, excluding them from holding status, noting “no rule can be the ratio decidendi from which the actual judgment . . . does not follow.” *Id.* at 45.

fact were potentially relevant rather than irrelevant to the outcome of a case. At a more global level, therefore, clarity is promoted by limiting holding status to an identifiable class of propositions.

Finally, providing holding status to potentially relevant propositions might advance the goal of candor, because it would reduce the risk that a judge would resolve other issues differently to ensure that resolution of the particular issue is credited as a holding. Suppose, for example, that Justice Powell cared a great deal about his conclusion that diversity was a compelling interest, but not so much about his narrow tailoring analysis. If it were clear that his compelling interest conclusion would count as dicta, he might have had an incentive to switch sides in the case and find that the Davis plan was narrowly tailored to the goal of improving diversity.

Evaluating potentially relevant propositions thus demands a difficult balance between constraint and the other factors. As a general matter, we believe that the constraint factor predominates in this context because allowing judges to create holdings on every issue that possibly could be relevant to resolution of a case would create many chances for opportunistic resolution of legal issues, and because the class of cases in which judges are likely to care sufficiently deeply about issues pointing in the opposite direction of their preferred case resolution as to encourage them to switch sides is quite small.²¹⁸ We do recognize, however, that the Supreme Court, given its unique role relative to the federal and state judiciaries on issues of federal law and its limited docket capacity, might afford itself more leeway in defining its own holdings more broadly than would ordinarily be the case in other courts.²¹⁹

4. *Structured Analysis*

The classification of potentially relevant propositions is a sufficiently close call that other considerations might rebut our presumptive conclusions. One such factor is the problem of structured analysis, which, we will argue, should be sufficient to move some nonsupportive propositions from the status of presumptive dicta to holding. We have already argued, however, that Justice Powell's conclusion that diversity is a compelling interest should count as dicta, even though the compelling interest analysis is part of the structured strict scrutiny test. Using the

²¹⁸ See Stearns, *supra* note 177, at 129-42 (demonstrating narrow class of such cases).

²¹⁹ See *supra* text accompanying note 350.

taxonomy of nonsupportive propositions that we offered in Table 1,²²⁰ we would draw the line between “structured but unordered, with continuing relevance” and “structured and ordered, with continuing relevance,” while recognizing that this line may at times be difficult to draw.

Where a test has a strong natural order, and the resolution of the first part of the test turns out to be a nonsupportive proposition, the constraint concern is diminished. In such cases, the judge might have been able to violate the natural order and thus avoid resolving issues that point away from the court’s eventual disposition, but when a judge does follow the natural order of a test, the concern that the judge is reaching out to resolve an issue about which the judge has particularly strong concerns is substantially diminished. Similarly, when an analysis does not involve a particular test but presents a logically ordered analysis, this too diminishes the concern regarding constraint. Certainly, where a court finds standing before ruling against a plaintiff or petitioner on the merits, the standing conclusion should count as a holding even though it is opposite the ultimate judgment.²²¹

Unordered or weakly ordered tests do not seem to us to furnish a sufficiently strong argument to convert propositions from presumptive dicta to holding. Once again, we recognize that there seems no ready metric for assessing how strongly tests are ordered or where to draw the dividing line between ordered and unordered tests, but given that virtually all tests are at least articulated in an order by the courts originating them,²²² we would be wary of claims that tests are generally strongly ordered so as to convert nonsupportive propositions into holding. Because narrow tailoring analysis could build upon an assumed rather than established compelling interest, however, it is not obvious that what may be the more intuitive ordering—compelling interest then narrow tailoring—is a sufficiently strong natural ordering in the sense in which we use that term.²²³

²²⁰ See *supra* text accompanying note 81.

²²¹ Declining to elevate such standing determinations to holding would threaten to undermine the development of standing doctrine. If announcements of standing determinations counted as holdings only when they supported the judgment, this might bias standing doctrine in a more liberal direction, at least if one assume that judges will sometimes liberalize standing to reach favorable outcomes on the merits, but that judges who grant standing and then deny relief on the merits are more likely to take a somewhat grudging view of the standing issue. Justice Scalia’s concurrence in the judgment in *Miller v. Albright*, 523 U.S. 420, (1999), provides an illustration of grudging conferral of standing to then deny relief on the merits. See *id.* at 454 n.1 (stating that while “[a]s an original matter, I would agree with Justice O’Connor that [the asserted] ground [for third party standing] is inadequate, . . . I do not read our cases as demanding so significant an impairment of the rightholder’s ability to sue as she does,” and then proceeding to deny relief).

²²² The mere articulation of a test taking the form, to achieve result X, the court must find A, B, and C, should not be presumed to imply that A must precede B, which must precede C.

²²³ See *supra* notes 76-77 and accompanying text.

We recognize, however, that dominant lawyerly and judicial intuition might sometimes diverge from formal analysis on how to treat a particular test. If most lawyers and judges would find a particular test to be strongly ordered that a more formal analysis reveals to be weakly ordered, perhaps the dominant understanding should control. For example, while we have argued that strict scrutiny is a weakly ordered test, and that Justice Powell's nonsupportive assertion that racial diversity is a compelling state interest is therefore dicta, one might instead conclude that the proposition is a holding if lawyers and judges in general consider strict scrutiny to be a strongly ordered test. The dominant perception of lawyers and judges on the question of ordering provides an independent check against articulating holdings that are motivated primarily by the desire to resolve questions of legal policy rather than dictated by the need to resolve the immediate case. A nonsupportive proposition may therefore be classified as holding based upon a determination that it derives from the application of a test that either is strongly ordered or that is generally understood to be strongly ordered.

The problem of structured analysis also may provide a basis for converting to dicta some alternative justifications that presumptively count are holdings.²²⁴ The paradigmatic case, once again, involves standing, except this time the court finds that a litigant who lacks standing would have lost in any event. Even though the court's assessment of the merits provides an alternative justification for the judgment, it was improper for the court to reach the merits given that the standing denial is jurisdictional. When a court *does* consider the merits in such a case (as in a case in which jurisdiction is absent), that may suggest that the judge has a strong normative interest in the resolution of the merits. The constraint factor thus counsels against treating the court's conclusion on the merits as holding. Similarly, discounting the merits resolution as dicta promotes proper consideration. One of the commonly expressed justifications for standing is to ensure adversity concerning those issues to be resolved.²²⁵ Allowing a court to set a precedent on the merits of a nonjusticiable case would undermine that justification.

We recognize that the factors do not point unambiguously in one direction, however. Allowing courts to resolve issues that the litigants have briefed might contribute to resolving ambiguities in law and thus promote clarity. Meanwhile, by denying precedential status to the

²²⁴ See *supra* Part II.A.2.b.i.

²²⁵ Richard Fallon, Jr., *Marbury and the Constitutional Mind: A Bicentennial Essay on the Wages of Doctrinal Tension*, 97 CAL. L. REV. 1, 33 (2003) (observing that "standing rules require concrete adversity between the parties").

merits resolution of a nonjusticiable case, a judge who is particularly interested in resolving the merits may have an incentive to find the case justiciable in order to achieve a holding on the merits, thus threatening both candor and constraint. The observation that judges might manipulate justiciability doctrine is, of course, not new,²²⁶ and the judges most inclined to engage in such manipulation are likely those with particularly strong normative views on the merits resolution for which a more dispassionate standing determination might pose a barrier.

Where analysis is as strongly ordered as the questions of standing and the merits, however, that is presumably because the courts have concluded that there is a strong reason to resolve the second inquiry only after the courts achieves a certain resolution on the first. In such circumstances, where the resolution of the first inquiry makes the second inquiry irrelevant, we believe that resolution of the second inquiry should count as dicta. We would reach the opposite conclusion, however, with tests that are only weakly ordered, such as the strict scrutiny test. Evaluation of the second prong of a weakly ordered test despite the resolution of the test as a whole through the first prong is not so unusual as to raise red flags about a judge's motivation in doing so.

A variation on the problem occurs where a court skips the first prong of an ordered test to resolve the second prong, most commonly because the second prong is the easier to resolve. For example, a court might avoid an analytically difficult standing issue and conclude that the merits of the claim is frivolous. The Supreme Court has confronted cases in which lower federal courts have engaged in such "hypothetical jurisdiction." In *Steel Co. v. Citizens for a Better Environment*,²²⁷ Justice Scalia, writing for a majority on this issue, rejected the practice, claiming that it "offends fundamental principles of separation of powers."²²⁸ The Court, however, has not always insisted that courts follow the order of even relatively strongly ordered tests. Thus, in *Ruhrgas AG v. Marathon Oil Co.*,²²⁹ Justice Ginsburg, while claiming to adhere to *Steel Co.*, permitted a lower federal court to resolve a less difficult question of personal jurisdiction, which represents a waivable barrier to jurisdiction, without first resolving the more difficult and

²²⁶ Gene R. Nichol, Jr., *Rethinking Standing*, 72 CAL. L. REV. 68, 70 (1984) (asserting that "the Court has so severely manipulated the injury standard that the foundation of standing law is essentially incomprehensible"). For an article that takes a more sanguine view of the standing doctrine, and of the injury requirement in particular, see Stearns, *Standing Back from the Forest*, *supra* note 179.

²²⁷ 523 U.S. 83 (1998).

²²⁸ *Id.* at 94.

²²⁹ 526 U.S. 574 (1999).

nonwaivable question of subject matter jurisdiction, where the particular resolution of the easier jurisdictional question rendered moot the resolution of the more difficult one. The Court concluded that despite the difference in the power to waive these forms of jurisdiction, “there is no unyielding jurisdictional hierarchy.”²³⁰

Our point is not to resolve the difficult questions of hypothetical jurisdiction or inter-jurisdictional priorities. These doctrines do not concern whether the courts should consider the resolution of logically subsequent issues as precedent, but rather whether courts are permitted to skip steps in reaching them. Nonetheless, the factors that underlie our analysis suggest that these issues are closely related, and thus it is not surprising that the courts have made fine distinctions in this area. There are policy issues that counsel in favor of permitting courts to resolve cases based upon easy (even if subsequent) rather than difficult (even if prior) issues. Doing so allows for continued consideration of the difficult issue while providing adequate consideration of those issues that prove sufficient to resolve the case. Clarity might be furthered as the legal system benefits from incremental development of doctrine, which is promoted by limiting the obligation of courts to reach issues in resolving cases. And when the second issue is genuinely easier, a judge’s decision to resolve the case on that issue does not suggest that the judge had a particularly strong normative interest in that issue.

On the other hand, once judges are allowed to skip a preliminary issue, judges who particularly wish to resolve the second issue might present the subsequent issue as easier than it actually is, and might set out a simplistic analysis of that issue to reinforce the judge’s claim that skipping the logically prior issue is justified. In addition, there remains a concern that skipping the first issue will promote underdevelopment of doctrine in that area, especially if the area is viewed as an arcane body of jurisdictional case law. In the narrow class of cases in which it is permissible to skip jurisdictional predicate issues to get to the merits, the courts have presumably found that concerns about manipulation are not sufficient to force compliance with the general rule.

²³⁰ *Id.* at 577.

5. *Hypotheticals*

The problem of hypotheticals proves one of the easiest to resolve. Hypotheticals and analogical reasoning sometimes help to sharpen case issues. The power of reasoning by analogy might suffice to make the resolution of hypothetical issues a sufficient, or even necessary, proposition in a case from the perspective of the deciding court. Nonetheless, the power to create holdings through hypotheticals presents significant problems of constraint. If a court, by resolving a hypothetical case, can determine the result of that case should it arise, there will be a strong incentive for judges to pose hypotheticals that they are particularly interested in resolving. This might result in courts devoting excessive attention to, and sometimes purporting to resolve, potentially relevant hypotheticals even absent adequate briefing, thus undermining the goal of proper consideration. Assigning holding status to hypotheticals would diminish doctrinal clarity in the law, at least to the extent that abstract or tangential hypotheticals obscure what a judge was actually required to resolve in the immediate case. Providing holding status to hypotheticals would thus threaten to undermine candor, as judges would increasingly claim to identify hypotheticals as relevant to the immediate case when their real interest is just resolving the hypotheticals themselves. Thus, propositions that are hypothetical should count as dicta.

6. *Biconditional Statements*

Even a hypothetical, however, conceivably could count as a holding if it is the inverse of a holding and part of a biconditional statement. The case for counting both halves of a biconditional statement as a holding is strong. When the case reasoning directly results in an “if, then” holding on a particular issue, there is little reason to believe that the judge has reached out in resolving the inverse issue. A judge who concludes that the inverse proposition is true may reach that issue because the judge feels particularly strongly about the issue, but there is at least a strong possibility that the judge is simply trying to offer a comprehensive resolution to the relevant question presented. For example, in *Bakke*, Justice Powell resolved the question of how a university could use race as a factor in an admissions program in a way that would be narrowly tailored to advance the interest of diversity.

The concerns about constraint that ordinarily would arise with respect to nonsupportive propositions thus seem greatly reduced when the nonsupportive proposition is the inverse of a

holding and thus part of what the court intended as a biconditional test. The consideration factor and the clarity factor point in opposite directions, but we believe that the pull of the clarity factor is stronger. Delaying final resolution may always promote consideration, but there is no special reason to delay resolution of inverse propositions. *Bakke*, for example, unquestionably presented the question that Powell resolved, and waiting until additional cases raised the same issue would not have greatly improved the Court's ability to address the issue. The clarity factor, meanwhile, strongly suggests that inverse propositions generally should be holdings. If a court cannot issue biconditional holdings, it may take considerably longer to provide definitive guidance on where the line is. Of course, sometimes there will be reasons for judges to conclude that given the complexity and fact-dependent nature of an issue, the consideration factor will outweigh the clarity factor, but we do not believe that the holding-dicta distinction should demand that consideration trump clarity in every such case.

Finally, the concern for judicial candor is significant here, as the *Bakke* example shows. If Justice Powell had been clearly prohibited from treating his analysis of whether the Harvard plan was narrowly tailored as holding, and if he had feared that a holding limited to striking down the Davis program might signal that no race-based admissions programs could be narrowly tailored or at least have left open that possibility, then he might instead have elected to go further in deciding which programs to allow. A regime that flatly prohibits the issuance of biconditional holdings might have placed Powell in the difficult position of *either* striking the Davis plan on narrow tailoring grounds, and leaving open the possibility that a later Court would extend *Bakke* to cover the Harvard plan, *or* endorsing the Davis program, even though Justice Powell truly believes that it was not narrowly tailored.²³¹ By instead issuing a biconditional holding, Powell was able to issue an opinion that promoted judicial candor in the sense of informing the relevant institutions as to the precise, and nuanced, basis for his narrow tailoring holding. The holding-dicta distinction ideally should not induce judges to take disingenuous positions on issues, contrary to how they actually believe the issues should be resolved. This problem might also distinguish the Supreme Court given the relative rarity with which it reviews any particular

²³¹ Of course a similar analysis can be applied to the issue of whether diversity is a compelling interest. If Justice Powell believed that he had identified a means of narrowly tailoring a race based admissions program but to an end that would be rejected as not compelling, he might instead have elected to sustain the Davis program based upon one of the advanced grounds that he did not believe satisfied strict scrutiny. But since the compelling interest prong does not form part of a biconditional holding, *see supra* x, we nonetheless conclude that this part of the Powell opinion is dicta.

substantive issue. The hiatus in equal protection analysis of affirmative action programs in higher education admissions from 1978 when the Supreme Court decided *Bakke* until 2004 when the Court decided *Grutter* and *Gratz* provides a credible normative justification for affording that Court more flexibility to determine the scope of its own holdings as compared with courts that lack discretionary power to control their dockets.

As a general matter, we believe that the inverse statements of holdings generally should count as holdings as well, but there is a significant caveat. It often may not be obvious that an inverse of a holding in fact is an inverse of a holding, when the holding and the inverse statement are not conceptually connected in the form of a biconditional. Allowing all inverse statements to count as holdings thus makes the holding-dicta distinction itself difficult to apply, undermining the goal of clarity. We would thus limit our conclusion to situations in which a court clearly is creating a test that is equivalent to a biconditional statement. An important virtue of this approach is that we believe that it comes closer to judicial practice and lawyerly intuition than an approach that counts all inverse statements of holdings as holdings. Our anecdotal evidence suggests that lawyers generally believe that when a court creates a test, even a biconditional test, that is a holding, but lawyers do not generally have an intuition that all inverse propositions of holdings should be holdings. Later, we will consider a case presenting inverse propositions that should not count as holdings.²³²

We recognize that this approach introduces a concern with respect to the clarity factor, albeit one that we view as less significant: determining whether the court is crafting a test. Although we think the answer in *Bakke* is affirmative, that case nonetheless illustrates the difficulty. Powell could have stated more explicitly that he was announcing a distinction between quotas and plus factors as the test for meeting narrow tailoring. Even so, the true test of a theory may be whether it can identify which cases are difficult, and our analysis can show precisely why the discussion of the Harvard plan presents a difficult issue for the holding-dicta distinction.²³³ Some readers of *Bakke* may see the discussion as creating a test, and those readers will tend to think the discussion created a holding. Others may see the discussion as unconnected

²³² See *infra* note 412 and accompanying text.

²³³ An additional reason for the difficulty is that Powell expressed the discussion of the Harvard plan in the form of a hypothetical. But that should not obscure that the Harvard discussion was an example used to advance a proposition—“if plus factor, then narrowly tailored”—that, though counterfactual, was the inverse of a holding. That is the reason that we believe that this proposition should count as a holding despite its status as a hypothetical.

to Justice Powell's conclusion that the Davis plan was not narrowly connected, and those readers will tend to think the discussion was dicta. We side with the former readers because Powell did explain that the reason that he was discussing the Harvard plan was because it bore on the issue of the extent to which a university can use race.²³⁴ But readers need not accept our conclusion to agree with our more general position that biconditional statements that announce tests create holdings, while other biconditional statements do not.

7. *Breadth*

The holding-dicta distinction should not force judges to resolve cases in the narrowest possible manner that is consistent with the case judgment. When an issue is genuinely before a judge, and thus concerns about constraint are at their minimum, the holding-dicta distinction should leave the choice to the deciding jurist to weigh the benefits and costs of resolving the issue in a complete and comprehensive manner. We recognize that there may be times when it is appropriate for a judge to craft holdings that have very narrow application, leaving it to future courts to identify which features of the case were dispositive, but there may be other cases in which it is appropriate for a court to clarify the law by offering a broader holding.²³⁵

As before, there is some risk that allowing judges to determine the breadth of their holdings will lead to an issue's being resolved in a comprehensive way by the first judge who has strong feelings about an issue. But any judge will be constrained by the obligation to harmonize a case with the holdings of earlier cases. The norm that judges should not disingenuously distinguish prior cases itself provides a significant constraint, complementing the constraint that a judge can issue holdings only along the decisional path to the case result. What is decisive here, however, is not a conclusion that constraint is irrelevant. Rather, it is the observation that a regime prohibiting unnecessarily broad antecedents would, except in cases involving patently absurd distinctions, effectively be a regime in which cases are generally distinguishable on their facts, and in which generalizations by a court would rarely provide the basis for holdings. We will turn to the task of critiquing such a regime soon.²³⁶ For now it is sufficient to observe that this type of regime is in tension with the premise of this Article that

²³⁴ See *supra* note 64 and accompanying text.

²³⁵ For an analysis demonstrating that the Supreme Court's narrowest grounds rule implies a choice of breadth when a unanimous court or a majority agrees to the chosen level of breadth, see *infra* notes 316-318 and accompanying text.

²³⁶ See *infra* Part IV.A.1.

statements in opinions sometimes constitute holdings. It is always possible to make statements narrower and more dependent on the particular facts of a case, but our system of precedent sometimes counts generalizations beyond the facts of a case as holdings.

That does not mean, however, that *all* broad statements should be considered to be holdings. Imagine the following hypothetical statement in *Bakke*: “If a graduate admissions program uses a quota, or an undergraduate admissions program uses a quota or a plus factor, then the program fails narrow tailoring.” Allowing that proposition to be a holding would effectively provide an end run around our conclusion governing hypotheticals. No undergraduate admissions program was before the Court. It would be fine for a court to sweep graduate admissions and undergraduate admissions together—“if any university admissions program uses a quota, then it is fails narrow tailoring”—even though such a conclusion would resolve no greater a number of cases. But a court cannot invoke a narrow antecedent for the case at hand, and then resolve additional cases not fitting within that antecedent. Stated conceptually, a court can fit the facts of a case within a broad circle and resolve all the fact patterns within that circle, but it cannot then annex an additional circle and resolve the fact patterns within that circle too.

8. *Summary*

Before analyzing the definitions of holding offered by other scholars and setting out our own, we will summarize the preceding discussion. The following table presents each of the problems identified in Part II and evaluates them according to the four normative criteria developed in this part. This table repeats each of the problem categories from Table 3, indicating whether the category has potential relevance for presumptive holdings or presumptive dicta. The second column identifies factors that provide an argument for altering the presumptive status of propositions based on this category, and the third column identifies factors that provide contrary arguments. The fourth column offers a conclusion based on the analysis in the preceding subsections, indicating whether the problem category should convert presumptive holdings to dicta or presumptive dicta to holding.

Table 4. The problem categories and their resolutions

Problem category (and presumptive status of potentially affected statements)	Factors that argue for altering presumptive status	Factors that argue against altering presumptive status	Conclusions
Alternative possible justifications (applied to presumptive holdings)	The ability to choose a decisional path implies some freedom from constraint, as a judge may pick the path containing issues that the judge most wants to resolve.	(1) There are alternative possible justifications for a large number of propositions, so an exception here would limit clarity by preventing the development of law. (2) Disallowing reliance upon a single prong of a multi-prong test would undermine concerns for candor and consideration by encouraging resolution of issues unnecessary to resolving the immediate case.	Should not turn holdings into dicta.
Alternative justifications (applied to presumptive holdings)	Allowing judges to choose multiple decisional paths undermines constraint, as the judges who choose multiple paths may be particularly eager to resolve the relevant issues. The effect on constraint is minimal, however, assuming that judges already have the power to choose a path.	If alternative justifications were not holdings and alternative possible justifications also were not holdings, judges might have an incentive to reach a result on one issue opposite from that on the other to ensure that they create a holding, undermining the goal of candor.	Should not turn holdings into dicta.
Structured analysis (applied to presumptive holdings)	When a court reaches a second issue despite resolving a test on the first issue, it is reaching out to resolve the second issue, thus undermining constraint.	(1) Allowing courts to resolve briefed issues may advance clarity in the law, without reducing consideration. (2) Considering a second alternative justification to be dicta in some circumstances might lead a court to change its position on the first alternative justification.	Where a test is ordered, only the first of multiple alternative justifications should count as a holding. For example, resolution of the merits of the case after a finding of no standing is dicta.
Structured analysis (applied to presumptive dicta)	When the first prong of an ordered multi-part test produces a nonsupportive proposition, concerns about constraint are less than with most other nonsupportive propositions.	Allowing courts to resolve briefed issues may advance clarity in the law, without reducing consideration.	Where a test is ordered, the resolution of one prong should not be considered to be dicta merely as a result of resolution of a later prong.
Hypotheticals (applied to presumptive holdings)	If hypotheticals count as holdings, courts may have wide discretion to resolve issues not before them, undermining constraint.		Even if a hypothetical proposition is presumptively a holding because of its role in the argument, it should count as dicta.

Problem category (and presumptive status of potentially affected statements)	Factors that argue for altering presumptive status	Factors that argue against altering presumptive status	Conclusions
Biconditional statements (applied to presumptive dicta)	By allowing judges to hold on inessential issues, biconditionals threaten to undermine constraint. The effect on constraint is small, however, because a judge is still limited to issues along a chosen decisional path.	Allowing biconditionals can promote clarity and candor by avoiding a choice between extreme doctrinal positions.	The inverse proposition of a holding should count as a holding where the court effectively endorses both halves of a biconditional in a test.
Breadth (applied to presumptive holdings)	The judges who resolve cases broadly may tend to be those with strong normative views on the issues in those cases.	If broad statements are not allowed, all cases are subject to being distinguished on their facts, undermining clarity in the law.	A court is free to endorse broad propositions, as long as the propositions do not forge together distinct sets of factual issues in an effort to evade the characterization of hypotheticals as dicta.

IV. DEFINING HOLDING AND DICTA

Not surprisingly, given the central importance of delineating dicta and holding, there is a longstanding, and often insightful, literature that has focused on this question. In this part, we will begin by providing a critical overview of that literature. While our analysis will show agreement with many discrete insights, it will also underscore our larger point, namely that the failure of prior scholars to undertake a more comprehensive approach in setting out the problem of defining dicta, and specifically the failure to articulate categories of judicial assertions that give rise to disputes in characterization, has limited the efficacy of prior scholarly efforts in this important area.

While the analytical framework set out in the prior parts of this Article can improve conceptual clarity even if courts ultimately reject our definition, we believe that our definition has much to commend it both at a theoretical level and at the level of practical implementation. As with virtually any definition that one might offer, ours is prone to some difficult applications and to potential disputes in concrete case settings. But a definition that hubristically claimed to resolve all cases would undoubtedly mask serious difficulties and disagreements. Instead, our definition is sufficiently capacious to allow reasonable jurists in difficult cases to argue for different conclusions edified by the common theoretical foundation. Properly applied, our definition resolves many issues and reduces the set of substantial disagreements that will arise.

A. *The Inadequacy of Existing Definitions*

1. *Reconciliability*

We begin with Professor Michael Dorf's important article,²³⁷ first by evaluating his critique of a commonly held definition of holding (and by implication of dicta), and second by critiquing Dorf's alternative suggested definition.²³⁸ Professor Dorf's principal project is to attack the definition of holding as limited to a meaningful or plausible reconciliation of the facts and outcomes of prior cases. Under the facts-plus-outcome, or reconciliation, approach, a court's task is to identify a theory that can explain the results of previous cases, regardless of whether the precedent-setting courts themselves adopted the superimposed theory. We agree with Dorf's conclusion, though we disagree with some of his reasons and will offer some additional and, we believe, more persuasive reasons. His analysis focuses on the U.S. Supreme Court and its powers as an Article III court; in contrast, while we use several cases arising in Article III courts as illustrations, our analysis is not grounded in the requirements of Article III.²³⁹

Dorf uses case examples to illustrate in depth how the Supreme Court infamously reinvented the important line of cases concerning the President's authority to remove administrative agency officials.²⁴⁰ Perhaps the strongest indication of recasting arose in Chief Justice Rehnquist's decision in *Morrison v. Olson*,²⁴¹ in which a majority upheld the independent counsel provisions of the Ethics in Government Act of 1978, even though the Act prevented the President from exercising complete removal authority over the independent counsel. The independent counsel performs purely executive functions,²⁴² and a previous case had clearly provided that Congress cannot restrict the President's authority to remove officials performing purely executive functions.²⁴³ *Morrison* effectively supplanted what had been a clear line with a balancing test.²⁴⁴

²³⁷ Dorf, *supra* note x.

²³⁸ See *infra* Part IV.A.4.

²³⁹ This is an important point because the distinction between holding and dicta is general and emerged within non-Article III common law courts in addition to Article III courts. The theoretical foundation we have developed, see *supra* Part III, is comparably broad and does not depend on the particular limitations of Article III.

²⁴⁰ Dorf, *supra* note x, at 2009-24.

²⁴¹ 478 U.S. 654 (1988).

²⁴² *Id.* at 689-91 (acknowledging that the independent counsel performed purely executive functions, but finding that distinction no longer relevant).

²⁴³ *Humphrey's Executor v. United States*, 295 U.S. 602 (1935). As Dorf persuasively argues, see Dorf, *supra* note x, at 2008-20,

Using the euphemism of “present considered view” to describe the Court’s new position, Rehnquist rewrote longstanding doctrine by effectively ignoring broad positions in the prior cases to arrive at an otherwise unattainable result.²⁴⁵ Relying on this line of cases and others, Dorf argues that the Court has repeatedly invented new rationales and used “a too-narrow view of holdings . . . as a means by which judges evade precedents that cannot fairly be distinguished.”²⁴⁶ Dorf expresses the concern that constraining “holding” to facts-plus-outcome invites such manipulation given the ability to recast prior results in a new light.²⁴⁷ The removal cases certainly support this insight. Dorf then proceeds to offer a broader theoretical justification for rejecting the facts-plus-outcome approach.

Dorf persuasively argues that the reconciliation approach simply fails to provide a positive explanation of judicial practice: “[T]he courts do not accept the facts-plus-outcome view of holdings; when they are not busy circumventing precedent by abusing the holding/dictum distinction, judges typically pay a great deal of attention to the words as well as the results of judicial decisions.”²⁴⁸ Ironically, the fact that courts sometimes construct seemingly disingenuous distinctions, or ambitiously recast old doctrines in a new light, underscores Dorf’s argument that courts ordinarily evince an obligation to attend to and somehow reconcile previously articulated holdings. And yet, as Dorf acknowledges, the occasional practice (one that might be more frequent in the Supreme Court than elsewhere) of disingenuously distinguishing extant case law complicates the task of constructing normatively satisfying and functional definitions of holding and dicta.

Not all observers of the courts believe that precedents, or even the facts of cases, constrain courts. Some judicial politics scholars have used empirical methods to argue that

that case itself involved a disingenuously narrow reading of *Myers v. United States*, 272 U.S. 52 (1926).

²⁴⁴ *Morrison*, 478 U.S. at 691 (“[T]he real question is whether the removal restrictions are of such a nature that they impede the President’s ability to perform his constitutional duty, and the functions of the officials in question must be analyzed in that light.”). *Morrison* itself was narrowly read in *Edmond v. United States*, 520 U.S. 651 (1997), decided at around the same time Dorf’s article was published. See Nick Bravin, Note, *Is Morrison v. Olson Still Good Law? The Court’s New Appointments Clause Jurisprudence*, 98 COLUM. L. REV. 1103 (1998) (discussing *Edmond*). If the trend continues, *Edmond* will be read narrowly as well.

²⁴⁵ In dissent, Justice Scalia pointed up the revisionist nature of the majority analysis. See *Morrison*, 478 U.S. at 705 (Scalia, J., dissenting) (“Surprising to say, the Court appears to concede an affirmative answer to both questions [implicated by present doctrine], but seeks to avoid the inevitable conclusion that since the statute vests some purely executive power in a person who is not the President of the United States it is void.”).

²⁴⁶ Dorf, *supra* note x, at 1999.

²⁴⁷ *Id.* at 2039-40 (“To discard the rationale of an earlier decision without the kind of compelling reasons that justify any departure from precedent does more than merely reinterpret a past case.”).

²⁴⁸ *Id.* at 2037 (footnotes omitted).

precedent fails to explain appellate court decision making.²⁴⁹ One difficulty in proving or refuting such claims involves selection bias.²⁵⁰ Cases that fail to settle tend to present close issues,²⁵¹ and the Supreme Court's procedure for choosing most of its own docket weeds out many cases in which the relevant precedents provide relatively clear guidance, even if the lower courts have applied them incorrectly.²⁵² In addition, the claim to "stare indecisis,"²⁵³ meaning that precedent imposes no meaningful constraint on judicial decision making, rests upon heavy citation in support of, and in opposition to, both sides of a given issue in a case. The difficulty with relying upon such data to refute the constraining effect of precedent is that litigants are apt to cite cases concerning the most contentious issues, while issues well settled by precedent, and thus beyond credible dispute, fail to warrant heavy citation. In any event, the widespread practice of citing the reasoning of prior cases strongly suggests that precedent, and expressed reasoning within decided cases, matters. If it did not, then litigants and courts presumably would simply present persuasive arguments without bothering to structure their arguments in the context of precedent. This insight suggests that as a positive matter, both the facts and at least some of the generalizations of prior cases do affect the results in subsequent cases.

Of course, even if we are correct that courts do not limit the reach of precedents to facts plus outcomes, it remains possible that earlier generations of jurists afforded less deference than courts today to the express reasoning articulated in prior opinions. Consider, for example, the ancient distinction between the holding and the *ratio decidendi* (or more simply *ratio*) of a case. As Karl Llewellyn explained: The "actual *holding* must be stated quite narrowly," encompassing the "precise point at issue . . . that the case decided."²⁵⁴ The *ratio* provides the "generally

²⁴⁹ See, e.g., SAUL BRENNER & HAROLD J. SPAETH, *STARE INDECISIS: THE ALTERATION OF PRECEDENT ON THE SUPREME COURT, 1946-1992* (1995).

²⁵⁰ See Frank Cross, *Political Science and the New Legal Realism: A Case of Unfortunate Interdisciplinary Ignorance*, 92 NW. U. L. REV. 251, 289 (1997) (noting that attitudinal effects would likely to be most significant at the Supreme Court, because the Court is especially likely to hear cases that do not have a clear answer under the legal model).

²⁵¹ See George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1 (1984) (presenting a model in which the plaintiff wins half of cases that go to trial, regardless of the relative strength of plaintiffs' claims in the overall pool of cases).

²⁵² Thus, the Supreme Court criteria for granting certiorari indicate that error correction is not a sufficient basis for a grant. See SUP. CT. R. 10 ("A petition for a writ of certiorari is rarely granted when the asserted error consists of erroneous factual findings or the misapplication of a properly stated rule of law.").

²⁵³ BRENNER & SPAETH, *supra* note 249.

²⁵⁴ KARL LLEWELLYN, *THE CASE LAW SYSTEM IN AMERICA* 14 (Paul Gewirtz ed. & Michael Ansaldi trans., 1989). Llewellyn wrote the work in German in 1928 to 1929.

applicable rule of law on which the opinion says the holding rested.”²⁵⁵ The *ratio* must also be distinguished from the “‘the rule of the case,’ meaning the ‘principle’ for which a case stands,” as that principle may develop in subsequent case law.²⁵⁶ Llewellyn observed a similar contrast between *dictum* and *obiter dictum*.²⁵⁷ A dictum is a statement that is nonessential but may have “second-order precedential value.”²⁵⁸ Obiter dictum is “merely a remark in passing” and is thus entitled to no more weight than the expression “of a legal scholar.”²⁵⁹ Our point is not to resurrect seemingly dated terminology, but rather to show that even assuming that it was once clear, the distinction between holding and *ratio decidendi* has blurred,²⁶⁰ as has that between *dictum* and *obiter dictum*. While it is interesting to speculate as to why commentators in the early twentieth century offered more gradations of precedent than is common today,²⁶¹ our task is to provide a practical set of tools for distinguishing holding and dicta that can be employed to edify current practice.

Dorf offers two normative arguments against the facts-outcome approach. With respect, we find these arguments less persuasive than his observation that the courts treat holdings as

²⁵⁵ *Id.*

²⁵⁶ Llewellyn appears to see a case’s *ratio decidendi* as not controlling on later courts, or at least as more subject to revision than the holding. Llewellyn’s view may have been dated even in his own time, as at least some contemporaries of his saw things differently. *See id.* at 15. Consider, for example, the following explanation of a case’s *ratio*:

A precedent, therefore, is a judicial decision which contains in itself a principle. The underlying principle which thus forms its authoritative element is often termed the *ratio decidendi*. The concrete decision is binding between the parties to it, but it is the abstract *ratio decidendi* which alone has the force of law as regards the world at large.

JOHN SALMOND, JURISPRUDENCE 201 (7th ed. 1924).

²⁵⁷ A scholar active before Llewellyn recognized this distinction but questioned its utility:

Some authorities attempt to distinguish between dicta and obiter dicta, saying that if the court naturally, though unnecessarily, uses words by way of illustrating or limiting the doctrine necessary to the decision, those unnecessary words are dicta, but that if the court goes still farther out of its way, leaves the question that is actually under discussion, and for the sake of illustration or for some other reason discusses subjects wholly foreign to the case, the words thus dropped outside the natural pathway are to be called not simply dicta but obiter dicta. . . . For practical purposes the expressions are interchangeable.

WAMBAUGH, *supra* note x, § 13, at 18-19.

²⁵⁸ LLEWELLYN, *supra* note x, at 14.

²⁵⁹ *Id.*

²⁶⁰ *See, e.g.,* Press-Enterprise Co. v. Superior Court, 478 U.S. 1, 28 (1986) (Stevens, J., dissenting) (“Despite the Court’s valiant attempt to limit the logic of its holding, the ratio decidendi of today’s decision knows no bounds.”); Johnson v. United States, 333 U.S. 46, 56 (1948) (Frankfurter, J., dissenting in part) (using the words “holding” and “*ratio decidendi*” separately, but making no apparent distinction between them); Harkless v. Sweeny Ind. Sch. Dist., 427 F.2d 319, 321 (5th Cir. 1970) (using “ratio decidendi” and “holding” apparently interchangeably); Krupnick v. Hartford Accident & Indem. Co., 28 Cal. App. 4th 185, 199 (1994) (finding “ratio decidendi” synonymous with “holding”). Indeed, even his contemporaries noted the increased attention that Llewellyn gave over time to what he labeled the *ratio*. *See* Herman Oliphant, *A Return to Stare Decisis*, 14 A.B.A. J. 71, 159 (1928).

²⁶¹ Maybe the reason is that in an era of formalism, any seeming departure from a pure holding had to somehow be justified. The labels provided a doctrinal mechanism of placing such extensions in boxes that appeared to preserve the holding-dicta line. In our post-realist world, we have no need to do that since it is a routine observation that courts manipulate these matters. The question is what are the limits on manipulation, and a viable framework for distinguishing holding and dicta will help to explain that.

significant, at least when not evading them. First, while acknowledging Justice Holmes’s assertion that “the common law evolves through a process of rationalization,”²⁶² Dorf insists that “reconceptualization has its limits” and that the facts-outcome view is “ultimately incoherent.”²⁶³ The problem, Dorf argues, is that “no mechanical rules can be devised to determine the level of generality intended by the precedent court.”²⁶⁴ We agree that there can be no mechanical rules for determining precedents in a facts-outcome regime, but we disagree that this makes the reconciliation view “incoherent.” Indeed, presumably the strongest argument for the facts-outcome approach is that its flexibility frees courts from strict adherence to the reasoning of their predecessors, thus allowing them to refine doctrine within a context of overall judicial constraint. A jurisprudential system providing no precedential weight to written opinions might be criticized for producing too little clarity,²⁶⁵ but incoherence arises only from mutually inconsistent constraints, not from indeterminacy.

Second, Dorf argues that “reconceptualizations rarely pay sufficient attention to the real issues at stake in the earlier case.”²⁶⁶ A court distinguishing an earlier case may do so on the basis of a fact to which the litigants in that case paid little attention.²⁶⁷ “When a court attempts to re-explain a prior decision,” Dorf argues, “the court necessarily focuses greater attention on the case before it at the moment than on the prior case.”²⁶⁸ Dorf does not explain, however, why this admitted tendency to read old cases in a new light is undesirable or, more generally, why the legal system should systematically prefer the original court’s articulated rationale to an arguably consistent alternative rationale—one that facilitates a desired judicial outcome—provided that it can be meaningfully constructed from the prior case. The original court, after all, might have selected its rationale without fully anticipating the implications of its immediate holding for a

²⁶² Dorf, *supra* note x, at 2035 & n.138 (citing OLIVER W. HOLMES, THE COMMON LAW 5 (1881)).

²⁶³ *Id.* at 2036.

²⁶⁴ *Id.* (quoting MELVIN A. EISENBERG, THE NATURE OF THE COMMON LAW 53 (1988)).

²⁶⁵ The civil law system has no doctrine of stare decisis, but precedents may have some persuasive weight, and this weight may become stronger when multiple courts arrive at the same conclusion. See Vincy Fon & Francesco Parisi, Judicial Precedents in Civil Law Systems: A Dynamic Analysis (2004) (unpublished manuscript, available at http://ssrn.com/abstract_id=534504) (providing an economic analysis comparing the civil law’s approach to precedent to that of the common law).

²⁶⁶ Dorf, *supra* note x, at 2038.

²⁶⁷ Dorf offers *Humphrey’s Executor* as an example: “How did the *Humphrey’s Executor* Court know with such certainty that a postmaster carries out inherently executive functions? Might not the setting of postage rates be considered quasi-legislative? Could not the application of those rates to particular parcels be termed quasi-adjudicatory?” *Id.* These arguments attack *Humphrey’s Executor* for positing facts that may not have existed in *Myers*. Yet one can imagine a definition of holding that permits distinctions, but only where the distinctions are clearly supported by facts deemed material by the original court.

²⁶⁸ *Id.*

significant future case. The second judge is in a better position to consider the operative precedent in light of at least two sets of facts: those motivating the initial precedent, and those in the immediate case. Thus, Dorf's argument might even cut in favor of the reconciliation approach.

Larry Alexander has also offered an argument against the reconciliation approach,²⁶⁹ which he labels "the result model of precedent."²⁷⁰ After distinguishing various formulations of this model, Alexander argues that they are all ultimately equivalent to Ronald Dworkin's general jurisprudential approach, demanding the articulation of a morally sensible principle that encompasses the relevant body of precedent and reconciles it with the immediate case.²⁷¹ Under this approach, a court must identify an admittedly imperfect but neutral governing principle that best harmonizes the cases.²⁷² Because a system of precedent implies reliance upon decisions with which an applying court might disagree as an original matter, the challenge under the reconciliation approach is for the court to reconcile those cases whose outcomes it prefers with those cases that it would have decided differently. So viewed, the defects of the facts plus outcome approach to precedent are the familiar defects of Dworkin's project.²⁷³

Professor Alexander provides a detailed attack on Dworkin, but for our purposes, it is sufficient to focus on one argument. Dworkin's task, Alexander notes, is to answer the following question: "'Which incorrect political/moral theory is 'best' from the point of view of correct political/moral theory?'"²⁷⁴ No theoretical apparatus "can answer such a question because the question is so bizarre," Alexander insists.²⁷⁵ We need not assess whether Alexander is correct in his assertion, because it is sufficient for our analysis to note that the task of identifying the best incorrect theory leaves considerable room for disagreement. The reconciliation approach might suffice in doctrinal areas with so few cases that one theory can explain them all. But the approach threatens to disintegrate as the underlying doctrinal area expands and gains sufficient complexity that it becomes implausible to offer a single theory that can credibly account for even

²⁶⁹ See Larry Alexander, *Constrained by Precedent*, 63 S. CAL. L. REV. 1 (1989).

²⁷⁰ *Id.* at 29-33.

²⁷¹ *Id.* at 33-34.

²⁷² See generally RONALD DWORKIN, *LAW'S EMPIRE* (1986) (providing an overview of Dworkin's jurisprudential approach).

²⁷³ For a critique of Dworkin's attempt to make the law coherent, see Ken Kress, *Why No Judge Should Be a Dworkinian Coherentist*, 77 TEX. L. REV. 1375 (1999).

²⁷⁴ Alexander, *supra* note x, at 39.

²⁷⁵ *Id.*

the vast bulk of cases. In many areas of law, commentators can agree only that the law is a “mess,”²⁷⁶ and while some commentators and jurists offer positive accounts that they trumpet as less messy than existing doctrine, or perhaps even neat, they generally will concede that even their new approach cannot explain all existing cases.²⁷⁷

But even if one could devise a comprehensive framework with which to embrace all or even most of the cases in a complex field, the reconciliation approach would remain normatively problematic. Perhaps ironically, it becomes problematic precisely because of the potential to develop such a comprehensive theory. Thus, in a system that truly cared only about a theory that could reconcile the facts and outcomes of decided cases, the law review article through which a professor best or most comprehensively reconceptualizes a seemingly problematic body of case law would—or at least should—have as much, and perhaps more, weight than admittedly partial judicial explanations. Even the most ardent law-and-economics positivists would not claim that their recapitulations *are* the law. Rather, the invariable claim is that such efforts *best explain* the law. Nor is this distinction semantic.²⁷⁸ Judges generally cast their opinions in the language of legal doctrine, rather than that of meta-theory, whether economics, moral philosophy, or some other methodology.²⁷⁹ And the incrementalism of doctrine necessarily imposes constraints that operate in tension with any theoretical claim to robustness. Moreover, academic claims to robustness are as contested as the doctrines themselves, and given the incentives of those proffering such explanations, they should be. Predictable doctrine in a reconciliation world would likely be impossible, at least in doctrinal areas involving complex and conflicting policy concerns.

²⁷⁶ See, e.g., John A. Ferejohn & Larry D. Kramer, *Independent Judges, Dependent Judiciary: Institutionalizing Judicial Restraint*, 77 N.Y.U. L. REV. 962, 1010 (2002) (calling the law of standing a “jumbled mess”); Marc R. Poirier, *The Virtue of Vagueness in Takings Doctrine*, 24 CARDOZO L. REV. 93 (2002) (“Everyone has heard the grumbling about the vagueness or messiness of the doctrine of regulatory takings.”); David A. Sklansky, *Traffic Stops, Minority Motorists, and the Future of the Fourth Amendment*, 3 1997 SUP. CT. REV. 271, 291-92 (citing various sources agreeing that Fourth Amendment law is a “mess”).

²⁷⁷ See, e.g., STEARNS, *supra* note x, at 251 (positing that “we cannot expect entirely consistent standing case law,” while presenting a theory intended to better harmonize existing standing case law).

²⁷⁸ For an interesting case illustrating this difference, consider *Lake River Corp. v. Carborundum Co.*, 769 F.2d 1284 (7th Cir. 1985), in which after providing a comprehensive analysis based upon law and economics criticizing an Illinois rule invalidating treating certain liquidated clauses as penalties, Judge Posner applies the Illinois rule to strike down such a clause.

²⁷⁹ For a defense of seeking agreement on doctrine rather than on broader principles, see Sunstein, *supra* note 194, at 1746-51.

2. *Factual Materiality*

The reconciliation approach, however, is not always treated as a license for courts to accept just any theory that reconciles the case law. Seventy years ago, Arthur Goodhart offered a variant on the facts-plus-outcome approach that he saw as producing some predictability.²⁸⁰ Parts of his analysis appear dated,²⁸¹ but because Goodhart furnishes an important early and often incisive effort to distinguish holding and dicta, his analysis deserves close attention.²⁸² Goodhart addresses some of the problems we have identified, and he offers important insights without providing a complete resolution.²⁸³ We begin with an algorithm that Goodhart claims can be used to identify the holding of a case, and thus to reduce the scope of uncertainty that underlies the facts-plus-outcome approach. In considering Goodhart's view, we implicitly assess the more famous, though less formally explicit, methodology of his contemporary Karl Llewellyn, who like Goodhart focuses on a consideration of the materiality of various facts.²⁸⁴

²⁸⁰ Goodhart, *supra* note 12.

²⁸¹ Goodhart carefully analyzes the precedential value of cases where courts have not issued opinions and where different reporters indicate different versions of the facts, small problems today. *See id.* at 169-70. Another limitation is Goodhart's failure to recognize the possibility that mixed questions of fact and law provide a category somewhere in between those of facts and law. *See, e.g., id.* at 171 (classifying as a question of fact what would now generally be regarded as a mixed question). Goodhart's assumption that the resolution of some mixed questions are simply fact findings prevents him from recognizing the possibility that the finding itself might have precedential value in other cases.

²⁸² Another reason for paying close attention is that some courts use an approach similar to Goodhart's. *See, e.g.,* Local 8599 v. Board of Education, 162 Cal. App. 3d 823, 834 (1984) ("The basic formula [for distinguishing holding from dicta] is to take account of facts treated by the judge as material and determine whether the contested opinion is based upon them.").

²⁸³ For example, Goodhart implicitly recognizes the problem of biconditional statements. *Id.* at 180 ("[I]f in a case the judge holds that a certain fact prevents a cause of action from arising, then his further finding that there would have been a cause of action except for this fact is an obiter dictum.").

²⁸⁴ *See, e.g.,* LLEWELLYN, *supra* note 217, at 47-50. Llewellyn also distinguishes between holding and dicta explicitly, but the distinction he offers is of little assistance:

When [a court] speaks to the question before it, it announces *law*, and if what it announces is new, it legislates, it *makes* the law. But when it speaks to any other question at all, it says mere words, which no man needs to follow. Are such words worthless? They are not. We know them as judicial *dicta*; when they are wholly off the point at issue we call them *obiter dicta*—words dropped along the road, wayside remarks.

Id. at 42. The question, of course, is how to determine when a question is properly considered before a court. While Llewellyn does offer additional insights on the holding-dicta distinction, we pay less attention to his analysis than to that of Goodhart because Goodhart has offered a clearer methodology. We recognize that Llewellyn's method is less subject to the objection below that Goodhart's method is wooden, but that is largely only because it is less precise. Llewellyn himself was skeptical of Goodhart's more mechanical approach. *See* KARL N. LLEWELLYN, *THE COMMON LAW TRADITION: DECIDING APPEALS* 14 n.9 (1960) (criticizing Goodhart for exalting the appellate opinion "into providing a Never-Never Single Answer," and dismissing Goodhart as "hopelessly off target," but not explaining further). We also disagree with the characterization that a court "legislates" when creating law in the course of resolving questions implicated in cases properly before it. In our view, judicial lawmaking is distinguished from legislating and is legitimated by the process through which it takes place, namely the need to resolve issues that arise in those cases through factors beyond the court's control. *See supra* Part III.A. *See also* STEARNS, *supra* note 38, at 198-202 (distinguishing legislative from adjudicatory lawmaking based upon the process through which each institution is called upon to make collective decisions).

Goodhart rejects the view that we can ignore a judge's reasoning altogether, noting that the judge necessarily makes law in his selection of the facts of the case to report.²⁸⁵ For Goodhart, "[i]t is by his choice of the material facts that the judge creates law,"²⁸⁶ and a holding is thus the proposition that given the set of material facts, the result of the case obtains.²⁸⁷ The challenge is thus to determine which case facts are or are not material. Goodhart recognizes judicial discretion in identifying facts as material, and suggests a resolution that takes the form of the following rebuttable presumption: "the facts of person, time, place, kind, and amount are presumably immaterial."²⁸⁸ Conversely, other facts are presumed to be material.²⁸⁹

In either case, the presumption is rebuttable. Thus, Goodhart asserts that the court can treat as immaterial facts that are generally presumed material,²⁹⁰ and can specify that facts generally presumed to be immaterial are material in the actual case.²⁹¹ In either instance, Goodhart claims, the court's statement of materiality or immateriality controls. Goodhart's analysis can perhaps best be conceived as an intermediate position between one that allows courts to articulate the *ratio decedendi* and one that instead limits the case to the facts plus outcome. While Goodhart would allow the judge to articulate a rationale that links the case facts to the judgment, the purpose of the announced rule is just that. It is a vehicle through which the judge, with some discretion in choosing among the various case facts, can weed out particular facts as immaterial.

Goodhart's approach has been subject to two related critiques. Julius Stone argued that facts may be identified at different levels of generality, and that different views about the generality of facts will produce different holdings.²⁹² Goodhart's method is therefore potentially indeterminate. And A.W.B. Simpson criticized Goodhart's distinction between a judge's identification of facts and a judge's statement of principles of law as wooden.²⁹³ The judge's

²⁸⁵ Goodhart, *supra* note 280, at 168-69.

²⁸⁶ *Id.* at 169.

²⁸⁷ *See id.* at 174-75 (giving examples of collection of facts and showing how different sets of material facts may make for broader or narrower holdings).

²⁸⁸ *Id.* at 174.

²⁸⁹ *Id.* at 178.

²⁹⁰ *Id.* at 175.

²⁹¹ *Id.* at 177.

²⁹² Julius Stone, *The Ratio of the Ratio Decedendi*, 22 MOD. L. REV. 597, 603 (1959).

²⁹³ *See* A.W.B. Simpson, *The Ratio Decedendi of a Case*, 20 MOD. L. REV. 413

enunciation of legal principles, Simpson argues, is what makes particular facts material,²⁹⁴ and, as a result, Goodhart's approach is equivalent to what Simpson refers to as the "classical proposition" that the *ratio* is the "principle of law which the judge considered necessary to the decision."²⁹⁵ As Simpson observes, this merely brings us back to where we started, since there is no single formula for identifying the *ratio decidendi*.²⁹⁶

In a response to his critics, Goodhart denied that accepting a judge's statements of legal rules (a commonly held view of precedent) was equivalent to Goodhart's preferred approach, deducing principles based on the material facts.²⁹⁷ And while Goodhart acknowledged the possibility of indeterminacy under his system, he insisted that no method could produce perfect determinacy.²⁹⁸ We agree with Goodhart on these two points, but we read his critics to make a broader claim. Stone and Simpson imply that ignoring judicial pronouncements of law in favor of the material facts that the judge has elected to report is an arbitrary method through which to identify holdings, and further, that the method is unlikely to achieve consensus even if everyone agreed to adopt it. Thus, while Goodhart's approach is systematic, it seeks to impose order without adherence to either the common practice of accepting a judge's articulated reasoning in achieving a decision or seeking an alternative normative justification with which to reconcile the prior result.

The weaknesses in Goodhart's model becomes manifest when Goodhart seeks to reach conclusions about specific issues. For example, in treating the issue of alternative holdings, Goodhart insists, "it is incorrect to say that either one of the conclusions involved a dictum

²⁹⁴ *Id.* at 414.

²⁹⁵ *Id.* at 413 (internal quotations and citation omitted). For an addition argument to the same effect, see J.L. Montrose, *Ratio Decidendi and the House of Lords*, 20 MOD. L. REV. 587, 592 (1957). We consider the possibility of defining holdings with respect to necessity below. See *infra* Part IV.A.3. Simpson's statement of the classical position adds a wrinkle, counting only what the judge *considered* necessary to the decision. Requiring an analysis of judicial intent, however, is unlikely to promote clarity in distinguishing holding from dicta. See *infra* notes 319-320 and accompanying text.

²⁹⁶ See A.W.B. Simpson, *The Ratio Decidendi of a Case and the Doctrine of Binding Precedent*, in OXFORD ESSAYS IN JURISPRUDENCE 148, 159 (A.G. Guest ed., 1961) ("There may indeed be as many ways of finding the *ratio* of a case as there are ways of finding a lost cat; certainly the *ratio* of some cases seems as elusive."). Despite Simpson's admonition, some computer scientists recently have tried to find algorithmic approaches to identifying a case's *ratio decidendi*. See, e.g., L. Karl Branting, *A Reduction-Graph Model of Ratio Decidendi*, in PROCEEDINGS OF THE FOURTH INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW 40 (1993), available at <http://citeseer.nj.nec.com/393344.html>. This literature considers the legal literature without questioning it, using it as a basis for identifying the task of the computer models.

²⁹⁷ Arthur H. Goodhart, *The Ratio Decidendi of a Case*, 22 MOD. L. REV. 117, 121 (1959).

²⁹⁸ *Id.* at 124 ("If it were possible to devise a method by which all precedents would become determinate, then the difficulties of interpretation would disappear, but I believe that this is a vain hope.").

because the one preceded the other or because the one was based on broad grounds and the other on narrow grounds.”²⁹⁹ While Goodhart does not explain this conclusion, he suggests that the judge who issues alternative holdings found both sets of facts to be material. Yet this observation alone cannot serve as a justification for according alternative holdings precedential status. The conceptual difficulty with this holding category arises because two sets of material facts redundantly lead to the same result. Goodhart’s methodology has no means of assessing the implications of redundancy.

We agree that Goodhart’s approach is excessively mechanical, or “wooden,” and that reliance upon reported material facts—even with the benefit of Goodhart’s rebuttable presumptions—cannot alone resolve the many difficult questions that arise in defining holding and dicta. As our preceding analysis suggests, however, we also recognize that the identification of material facts provides an important early step in constructing such a set of working definitions. Goodhart’s central contribution is in recognizing that judges inevitably have some discretion in selecting the issues for which they will make law, and that discretion is constrained, although not eliminated, by the facts presented. The difficulty, which lies at the core of the various critiques we have reviewed, is that facts, material or otherwise, do not speak for themselves. A system of rebuttable presumptions concerning potentially material and immaterial facts cannot form the framework for identifying either the governing theory of a case or the level at which, whichever theory is chosen, it should be expressed. Judges do not merely identify material facts, but they develop legal reasoning that applies the facts and that eventually reaches a particular result. A satisfactory definition of holding and dicta must therefore examine the reasoning that connects the material facts to the result, rather than relying solely upon the selected material facts.

3. *Necessity*

The most influential definition, perhaps largely because of its inclusion in *Black’s Law Dictionary*,³⁰⁰ identifies “dictum” as a statement in a judicial opinion that is “unnecessary” to the

²⁹⁹ Goodhart, *supra* note 280, at 180.

³⁰⁰ *Black’s* provides “*obiter dictum*” as an alternative definition for “dictum.” BLACK’S LAW DICTIONARY 1102 (Bryan A. Garner ed., 8th ed. 2004). “*Obiter dictum*” is then defined as “[a] judicial comment made while delivering a judicial opinion, but one that is unnecessary to the decision in the case and therefore not precedential (although it may be persuasive).” *Id.*; *supra* text accompanying note 13.

case resolution. The focus on whether a statement is necessary dates at least to Eugene Wambaugh, who explained, “So far as the opinion goes beyond a statement of the proposition of law necessarily involved in the case, the words contained in the opinion, whether they be right or wrong, are not authority of the highest order, but are merely words spoken, *dicta*, *obiter*, or *obiter dicta*.”³⁰¹ Perhaps the main virtue of Wambaugh’s test is its simplicity. As Wambaugh asserted, the test allows “even the beginner [to] determine whether it is possible for a given proposition of law to be involved in a given case”³⁰²:

In order to make the test, let him first frame carefully the supposed proposition of law. Let him then insert in the proposition a word reversing its meaning. Let him then inquire whether, if the court had conceived this new proposition to be good, and had had it in mind, the decision could have been the same.³⁰³

Wambaugh thus establishes a high threshold for necessity: If the decision *could* have been the same with the negation of a proposition, then the proposition was not necessary to the disposition and thus counts as dicta. In this analysis, sufficient but not necessary propositions are dicta.

While frequently cited by courts, Wambaugh’s approach (and along with it, that of *Black’s Law Dictionary*) proves the easiest to falsify. For that reason, we should not be surprised that it is inconsistent with standard judicial practices, including that of according precedential status to alternative justifications.³⁰⁴ Judge Kozinski noted the problem of alternative justifications in his disagreement with Judge Tashima in a recent en banc Ninth Circuit case. Judge Tashima invoked the “necessary” definition of holding,³⁰⁵ and Judge Kozinski pointed out the inconsistency:

Under his rationale, which is that everything not necessary to the result is dicta, both alternative holdings are dicta because neither is necessary to the result. We can test this proposition by asking the question: Would the result change if either of the alternative holdings were removed? The answer, of course, is no. Since

³⁰¹ WAMBAUGH, *supra* note x, § 13, at 18-19.

³⁰² *Id.* § 11, at 17.

³⁰³ *Id.*

³⁰⁴ See *Woods v. Interstate Realty Co.*, 337 U.S. 535, 537 (1940) (“[W]here a decision rests on two or more grounds, none can be relegated to the category of obiter dictum.”); see also *In re Hearn*, 376 F.3d 447 (5th Cir. 2004); *Whetsel v. Network Property Services, LLC*, 246 F.3d 897, 903 (7th Cir. 2001).

³⁰⁵ *United States v. Johnson*, 256 F.3d 895, 920 (9th Cir. 2001) (en banc) (Tashima, J., concurring). As Judge Tashima noted, the Ninth Circuit had invoked the *Black’s Law Dictionary* definition previously. *Id.* (citing *Export Group v. Reef Indus., Inc.*, 54 F.3d 1466, 1472 (9th Cir.1995)). Judge Tashima, however, did not consider whether that selection of a definition was itself a dictum.

either could be removed without affecting the result, neither is necessary, and so under Judge Tashima's reasoning, dicta.³⁰⁶

Writing for the Court, Judge Kozinski instead adopted an earlier case definition that, in the spirit of legal fictions, redefined “necessarily” to mean “only that the court undeniably decided the issue, not that it was unavoidable for it do so.”³⁰⁷

Judge Kozinski’s alternative appears untenable or at least incomplete.³⁰⁸ While Kozinski rescues *Black*’s use of “necessary” in defining “dicta,” he does so as an exercise in creative redefinition. Wambaugh’s test, in contrast, was based upon the actual meaning of “necessary,” and the court’s attempt to escape from Wambaugh’s approach reflects the deficiency of a necessity-based definition of holding. Wambaugh himself maintained that alternative justifications were dicta,³⁰⁹ but that the existence of alternative possible justifications was irrelevant.³¹⁰ The latter conclusion appears to be in tension with Wambaugh’s test, however, because whenever there is an alternative possible justification, it is at least possible that the court *could* have reached the same result even if it had negated the conclusion on the alternative issue that it did explore.

It is possible to devise a test that generally requires necessity as a precondition to attaching the status of holding, but that exempts alternative rationales from this general requirement to avoid the anomaly of concluding that the court has expressed no holding in such a

³⁰⁶ *Id.* at 915 (Kozinski, J., concurring). The same analysis applies when a court grants standing and then denies relief on the merits or when a court finds harmless error. The “holding,” which grants standing or finds error, can be reversed without affecting the case judgment, namely denying relief.

³⁰⁷ *Id.* (citing *United States v. Weems*, 49 F.3d 528, 532 (9th Cir.1995)). Of course this is no more fictitious than Chief Justice Marshall’s earlier effort to define “necessary” to mean “convenient, or useful, or essential.” *M’Culloch v. Maryland*, 17 U.S. (4 Wheat. 316, 413 (1819)). Judge Tashima’s definition is akin to convenient or useful in the sense of being “conducive to” the ultimate case disposition.

³⁰⁸ Judge Kozinski attempted to rescue it with the following clarification:

Of course, not every statement of law in every opinion is binding on later panels. Where it is clear that a statement is made casually and without analysis, where the statement is uttered in passing without due consideration of the alternatives, or where it is merely a prelude to another legal issue that commands the panel’s full attention, it may be appropriate to revisit the issue in a later case.

The problem with this approach is that it offers no attempt to constrain judges’ power to render decisions on any issues they please, as long as they are not casual or unanalytical in doing so. *See supra* Part III.C.1 (discussing the importance of constraint).

³⁰⁹ The endorsement is somewhat equivocal:

If the court finds error as to two or more points, the case cannot be an authority of the strongest sort upon any of them, if it is impossible to say that the court considered any particular one of them conclusive. Yet as to each of the errors pointed out by the court the decision certainly does have some authority.

WAMBAUGH, *supra* note x, § 25, at 33. Wambaugh does not explain how one might determine whether “the court considered any particular one of them conclusive.” Applying his own test, neither would be conclusive, as neither would necessarily be determinative.

³¹⁰ *Id.* (“If, however, the opinion definitely says that the reversal rests wholly upon one point, the case does become a precedent as to that point.”).

case. But even if we can thereby avoid a particular problem associated with the necessity requirement, we still need a normative justification for necessity in the first place. The failure of the necessity approach to accord with a fundamental practice in distinguishing holdings from dicta reveals that the approach does not match our intuitions about the holding-dicta line. Thus, even setting aside alternative justifications and alternative possible justifications, consider the case of an unnecessarily broad holding that is clearly embraced within an announced opinion. Any time a case could be resolved on narrower grounds, the broader rationale can be claimed unnecessary. Much judicial reasoning is unnecessary in the sense that a court could have resolved a case on narrower grounds, for example by issuing a fact-specific ruling without making the legal basis of a decision clear, or simply by declining to announce a broad rule when a narrower alternative is available.³¹¹

We have already concluded as a normative matter that judges should be allowed to create holdings broader than necessary, but it is worth also showing that judges are, in fact, allowed to do so in practice. To illustrate, consider the *Hopwood* decision, in which a split panel of the Fifth Circuit struck down the University of Texas Law School affirmative action program.³¹² Two panel members claimed broadly that race could not be used in furtherance of diversity, and that Justice Powell's *Bakke* opinion was not controlling. Concurring in the judgment, Judge Wiener maintained instead that the challenged program was invalid even under *Bakke* and that the majority's approach was improperly broad and unnecessary to achieving the case result.³¹³ Our point, of course, is not to defend the majority decision in *Hopwood*.³¹⁴ Rather, it is that *whether or not* the majority panel correctly interpreted *Bakke*, there is no question what it held. The court's holding was not the narrower position embraced by the concurrence. Instead, the holding was the broader and unnecessary view—unnecessary in the sense that the court could have achieved the same result with a narrower view that did not call the precedential status of Powell's *Bakke* opinion into question—embraced by the majority to resolve the case.³¹⁵

³¹¹ This harkens back to the difficulty with Goodhart's effort to rely upon selected material facts as the basis for identifying a holding; even if one agrees on which facts are material, this fails to resolve the level of generality at which the holding of the case should be cast. *See supra* text accompanying note 292.

³¹² *Hopwood v. Texas*, 78 F.3d 932 (5th Cir. 1996).

³¹³ *Id.* at 963, 967-68 (Wiener, J., specially concurring).

³¹⁴ Indeed, as we have made clear, we *do* believe that Powell's *Bakke* opinion expressed the holding of the Supreme Court, and thus was binding on the lower courts. *See supra* note 38.

³¹⁵ In dissenting from denial of rehearing en banc, Chief Judge Politz complained, "The majority of the panel overruled *Bakke*, wrote far too broadly, and spoke a plethora of unfortunate dicta." *Hopwood v. Texas*, 84 F.3d 720, 724 (Politz, C.J., dissenting).

We know of no rule that suggests that the holding of a split appeals court panel is expressed in that opinion that resolves the case on the narrowest grounds when the remaining two jurists agree to an alternative, albeit broader, rationale. Indeed, the narrowest ground rule, announced in *Marks v. United States*,³¹⁶ implies just the opposite. The rule states: “When a fragmented Court decides a case and no single rationale explaining the result enjoys the assent of five Justices, the holding of the Court may be viewed as that position taken by those Members who concurred in the judgment on the narrowest grounds.”³¹⁷ The negative pregnant of the announced rule is that when the Supreme Court is not fragmented (meaning that it issues a unanimous or majority decision), and thus when a majority embraces a single governing rationale, the holding is expressed in the majority opinion. Critically, this is so *even if* a justice concurring in the judgment would have preferred instead to resolve the case on a narrower ground. Extending this rule to the three-judge circuit court panel, the *Hopwood* majority, despite its broader holding and despite the narrower special concurrence,³¹⁸ stated the holding.

Applying Wambaugh’s approach to cases like *Hopwood* would potentially make a court’s decision to address a case on relatively broad grounds irrelevant to the precedential effect of the announced rule when a narrower alternative can be imagined. As suggested above, the *Marks* rule would then have to be adapted to apply even to majority decisions when a judge concurring in the judgment offers a narrower means of achieving the same result. Indeed, the implications might be broader still. The availability of an unarticulated narrower ground (as might occur, for example, in a case decided by a unanimous court) would potentially prevent the broader articulated ground from having precedential status. Thus, a law review article suggesting a narrower rationale that could have been used to achieve the outcome in a Supreme Court case

from denial of rehearing en banc). Politz’s only specific claim about dicta, however, was as follows: “This ‘injunction’ is wholly unnecessary to the disposition of the matter appealed and thus is clearly dictum . . .” *Id.* at 722 (Politz, C.J., dissenting from failure to grant rehearing en banc). At first glance, this language appears to ignore the distinction between the reasoning of a case and its disposition. It is the reasoning of the case, rather than the disposition, that must be classified as holding or dicta. The looseness of language here, however, lies in the word “injunction,” as the panel in fact denied an injunction. *See Hopwood v. Texas*, 78 F.3d 932, 958-59 (5th Cir. 1996). Chief Judge Politz’s complaint is that “two members of the three-judge panel determined to bar *any* consideration of race in the Law School’s admission process,” 84 F.3d at 722, and the word “injunction” reflects, however imprecisely, that the decision would have sweeping effects. Nonetheless, Politz makes no clear argument that the fact of sweeping effects is sufficient to classify any of the panel’s statements as dicta.

³¹⁶ 430 U.S. 188 (1977).

³¹⁷ *Id.* at 194 (internal quotation marks omitted).

³¹⁸ The circuit court uses “special concurrence” in the same manner as the Supreme Court uses concurrence in the judgment. Indeed, at the end of Judge Wiener’s special concurrence, he states, “Thus I concur in the judgment of the panel opinion but, as to its conclusion on the issue of strict scrutiny and its gloss on the order of the remand, I disagree for the reasons I have stated and therefore concur specially.” 78 F.3d. at 968 (Wiener, J., specially concurring).

could in theory state the law, rather than the opinion under review. A narrower ground suggested in an opinion concurring in the judgment would then prove a sufficient, but not a necessary, condition for disregarding the broader rationale embraced in a majority opinion.

Taken to its logical conclusion, this understanding of necessity would call into serious question twin premises of legal realism, first that judges make law, and second, that they have discretion in doing so. Ultimately, the analysis suggests that the necessity approach is no more than the reconciliation approach in a disguised form. For a proposition to be necessary to a judgment, there must be a unique chain of reasoning from the facts to the judgment dependent upon that proposition, or if there are multiple possible chains of reasoning, then each of the alternative chains must also be dependent upon that proposition. Thus, when a proposition is necessary to a judgment, there is no way to reach the same judgment with respect to the relevant facts without including the claimed proposition. Suppose, for example, that after an opinion containing a proposition that satisfies the necessity test is announced in a given case, another court would prefer to resolve a later case in a manner that depends upon the negation of the established proposition from the earlier decision. Assuming adherence to precedent, the second court would then be unable to resolve the subsequent case in the preferred manner while harmonizing the two cases. Thus, a proposition can be necessary to a judgment only if that proposition would also be controlling in a reconciliation-based precedent regime.

Wambaugh clarifies his test by providing that “a case is not a precedent for any proposition that was neither consciously nor unconsciously in the mind of the court.”³¹⁹ If we adopt this interpretation of necessity, then it becomes even more difficult to establish a precedent under the necessity regime than under the reconciliation regime.³²⁰ Even if a judge logically could not have reached the case outcome without endorsing a particular proposition, that proposition still would not be considered a holding if the judge did not at least subconsciously consider that proposition. The difficulty of determining subconscious processing aside, Wambaugh’s test would result in creation of a holding only when a judge *must have adopted* a proposition to arrive at the case outcome and in fact *did* adopt and articulate that proposition. Although this approach, unlike the reconciliation approach, pays some attention to a court’s

³¹⁹ WAMBAUGH, *supra* note x, § 17, at 24.

³²⁰ The “mind of the court” approach is reminiscent of the “meeting of the minds” concept in contract law. Both are fictions that are intended to capture objective indicia of the approach the court necessarily took, or that the parties agreed upon, respectively.

articulated chain of reasoning, it cannot be squared with any approach that credits the discretionary choices judges actually make in resolving cases.

4. Preclusivity

In developing his test for determining which portion of a court's articulated reasoning should be credited as holding, and which portions should be discounted as dicta, Professor Dorf imports the doctrine of issue preclusion.³²¹ In particular, Dorf focuses on a provision from the *Restatement (Second) of Judgments*: "When an issue of fact or law is actually litigated and determined by a valid and final judgment, and the determination is essential to the judgment, the determination is conclusive in a subsequent action between the parties, whether on the same or a different claim."³²² Dorf also relies upon a comment to this section that endorses the converse proposition. Specifically, when the articulated criteria for issue preclusion are not met, the determination is not deemed to have been resolved conclusively in subsequent litigation between the parties on the same or on a different claim.³²³ The explanatory comment explains that determinations that do not satisfy the criteria for issue preclusion are not deemed to have been resolved conclusively because "such determinations have the characteristics of dicta."³²⁴

While we agree that issue preclusion provides the basis for insights into distinguishing holding and dicta, Dorf fails to provide a foundation for relying upon preclusion law to develop his definition of holding. Dorf does note that preclusion law requires the applying court to distinguish more important or more central rationales from less important or less central rationales. That may suffice to show that preclusion law sheds some light on the holding-dicta analysis, but it does not establish the preclusion law analogy is sufficiently strong as to provide a sound normative basis for the definitions of holding and dicta. While the sets of concerns that these two bodies of law address overlap, they remain critically distinct.

³²¹ Dorf does not wholeheartedly endorse importing preclusivity. Instead, he states, "Perhaps we can gain some insight from a related field: preclusion law." Dorf, *supra* note x, at 2041. Earlier in the article, Dorf describes the portion discussing preclusion law as "explain[ing] what it means to say that a proposition is essential to the rationale of a case." *Id.*

³²² RESTATEMENT (SECOND) JUDGMENTS § 27 (1982), *quoted in* Dorf, *supra* note x, at 2041.

³²³ *Id.* § 27 cmt. h, *quoted in* Dorf, *supra* note x, at 2041 & n.159.

³²⁴ *Id.* Reliance upon the quoted provisions is potentially circular. Thus, Dorf relies upon issue preclusion to define holding and dicta because the relevant provision in the *Restatement* helps to explain what counts as "essential" within an opinion; meanwhile the same provisions base their discussion of what is essential, as needed to determine issue preclusion, upon the distinction between holding and dicta.

Preclusion law seeks to provide rules for which issues are deemed conclusive to the parties in future cases; it does not aspire to provide rules for which issues will bind nonparties. There are sound normative justifications for defining issues in the preclusion context more broadly than governing rationales in the context of defining holding. One of the policy justifications for preclusion is to encourage parties to air their legal disputes in a cost-effective manner, one that minimizes the burdens on judicial resources.³²⁵ Potentially redundant claims should therefore be minimized, and one way to achieve that is to construe “issue” sufficiently broadly that parties are encouraged by issue preclusion to present closely related claims in one proceeding.

At the same time, the normative considerations that are central to the holding-dicta distinction are reduced in the preclusion context. Preclusion doctrine, for example, is not particularly concerned with judicial incentives to reach out and resolve issues not closely tied to the material facts of a case. It may be worthwhile for appellate judges to anticipate issues that will arise immediately on remand, and there is thus an argument that law-of-the-case doctrine should be broader in its definition of holding than the holding-dicta distinction that we have been considering.³²⁶ Given that Dorf’s primary project was to criticize excessively narrow definitions of holding, it is not surprising that he carefully considered preclusion doctrine in his analysis. Our concern, however, is that in doing so, Dorf might have been led into endorsing an indefensibly broad definition of holding.

In considering the question of alternative justifications, for example, Dorf cites the *Restatement* for the proposition that “when an appellate court bases a decision on two grounds, each of which, standing alone, would support the judgment, preclusive effect will be given to both determinations.”³²⁷ He then posits that this alternative judgment rule from preclusion law

³²⁵ RESTATEMENT (SECOND) JUDGMENTS at Introduction (“Indefinite continuation of a dispute is a social burden. It consumes time and energy that may be put to other use, not only of the parties but of the community as a whole.”); see also Robert G. Bone, *Rethinking the “Day in Court” Ideal and Nonparty Preclusion*, 67 N.Y.U. L. REV. 193, 196 (1992) (noting that “increasing concern about the social costs of litigation” explains recent tendencies to increase the scope of preclusion, but that “[c]ourts have hesitated to expand nonparty preclusion—the preclusion of persons who were not parties to the first lawsuit—even though current nonparty preclusion rules . . . tolerate extensive relitigation at substantial social cost”).

³²⁶ The U.S. Courts of Appeals have generally held that dicta is not law of the case. See, e.g., *Bull HN Info. Sys., Inc. v. Hutson*, 229 F.3d 321, 326 n.3 (1st Cir. 2000) (“The law of the case doctrine does not apply to dicta.”); *United States v. Rice*, 1996 WL 44452, *4 (10th Cir. 1996) (listing numerous additional citations from various circuits). We have seen, however, that conventional definitions of holding and dicta are ambiguous.

³²⁷ Dorf, *supra* note x, at 2042.

can effectively be extended to distinguish holding and dicta in such cases.³²⁸ While Dorf acknowledges that given alternative holdings in a single case, “neither rationale is essential to the Court’s conclusion,”³²⁹ he nonetheless argues in favor of treating both rationales as holdings because “[t]he Court gave careful consideration to both rationales” and because “a decision based on either ground would have been legitimate.”³³⁰ While we ultimately agree with Dorf’s conclusion on alternative justifications, we find his explanation less persuasive. Our analysis has emphasized that a central concern of the definition of dicta is to minimize judicial overreaching and to ensure legitimacy in the construction of actual holdings. Careful consideration, standing alone, cannot suffice to elevate a judicial assertion into a holding.³³¹ Otherwise courts could resolve any issue for which they hold strong views by presenting a well reasoned opinion concerning its resolution.

In other respects, issue preclusion is too narrow to serve as a basis for the holding-dicta distinction. Even twentieth-century expansions of issue preclusion, for example nonmutual offensive issue preclusion,³³² which allows a nonparty to a prior suit to preclude a party to that suit from relitigating an adversely resolved issue,³³³ will not permit someone who is neither a party nor in privity with a party in the prior suit to be bound by any issues that were determined in that suit.³³⁴ This limitation cannot logically be applied in the holding-dicta context, because the distinction between holding and dicta applies to both parties and nonparties. If the holding-dicta distinction were defined without reference to that criterion, much of issue preclusion doctrine would be redundant, at least in cases within the same jurisdiction as an initial decision. If findings in a lawsuit were conclusive against parties and nonparties alike, nothing would turn on

³²⁸ *Id.*

³²⁹ *Id.* at 2044.

³³⁰ *Id.*

³³¹ If it did, anything that a court carefully considers would count as holding rather than dicta. It is more plausible to imagine careful consideration as one among many elements in identification of holdings, but we do not think that such a balancing test is necessary. Rather, lack of clear judicial consideration matters because it tends to leave ambiguous the level of abstraction at which a court decided a case, making it harder for subsequent courts to classify a holding at a particular level. *See infra* Part IV.B.1 (noting that there may be ambiguities about the level of generality at which a court resolved a particular issue).

³³² *See Parklane Hosiery Co., Inc. v. Shore*, 439 U.S. 322, 331 (1979) (approving of the use of nonmutual offensive issue preclusion, where not unfair to the defendant).

³³³ *See generally Comment Note—Mutuality of Estoppel as Prerequisite of Availability of Doctrine of Collateral Estoppel to a Stranger to the Judgment*, 31 A.L.R.3D 1044 (2004) (detailing the decline of the mutuality rule, and the current state of the law).

³³⁴ *Id.* n.7 (“The party asserting issue preclusion must establish that . . . the party against whom issue preclusion is asserted was a party or in privity with a party to the prior adjudication.”).

whether a party actually litigated a lawsuit, an issue to which the *Restatement* not surprisingly pays considerable attention.³³⁵

Even if preclusion law were not significantly over- and under-broad (another way of saying that it provides the wrong set of rules) for evaluating the definitions of holding and dicta, it would nonetheless prove an inadequate guide because the legal doctrines it establishes are too ill defined to provide meaningful guidance. Recognizing this, Dorf suggests that preclusion law can provide guidance based upon examples if not from clear rules.³³⁶ One difficulty, however, as Dorf himself recognizes, is that the case law sometimes contradicts the *Restatement*, even on the simple problem of alternative justifications.³³⁷ Looking to a body of law that itself is murky does not seem an edifying approach for the courts to adopt in resolving the holding-dicta distinction when the concern is that courts currently draw this distinction too imprecisely.

Even where the *Restatement* directly anticipates interpretive problems, it does not always offer clear resolutions. For example, statement c of § 27 acknowledges, “One of the most difficult problems . . . is to delineate the issue on which litigation is, or is not, foreclosed by the prior judgment.”³³⁸ Delineating any given issue, that is determining which conclusions about a particular issue are conclusive, is as important in defining the holding as is determining whether a portion of a case contains a holding at all. Statement c merely suggests a balancing test in cases in which the issues are not identical from one case to the next. The elements of the balancing test are open-ended and, as is appropriate to preclusion law but not to evaluating holding versus dicta, are geared to situations in which the same parties are litigating both cases.³³⁹

While Dorf appears to recognize some of the inherent limitations of the *Restatement* in providing guidance for defining dicta and holding, he nonetheless relies upon the principle of essentiality, which underlies much of preclusion analysis. One example that Dorf considers anticipates what we have called the problem of structured analysis, and more specifically the

³³⁵ See RESTATEMENT (SECOND) JUDGMENTS § 27 (using “actually litigated” language).

³³⁶ Thus, Dorf observes, “Fortunately, preclusion law consists of more than abstract statements of principle. At least in its widely accepted canonical form, the *Restatement of Judgments*, preclusion law takes clearer shape from examples.” *Id.* at 2041-42.

³³⁷ *Id.* at 2042 n.160. Dorf points out that some cases do provide preclusive effect to alternative judgments of trial courts. *Id.* (citing *Dozier v. Ford Motor Co.*, 703 F.2d 1189, 1194 (D.C. Cir. 1983) and noting a “division of authority”).

³³⁸ RESTATEMENT (SECOND) JUDGMENTS § 27 cmt. c.

³³⁹ For example, the comment asks, “Could pretrial preparation and discovery relating to the matter presented in the first action reasonably be expected to have embraced the matter sought to be presented in the second?” *Id.* There are two problems with applying this to the holding-dicta distinction. First, there is no indication of what may “reasonably be expected.” Second, the focus on pretrial preparation makes far more sense when the same parties are litigating an issue.

problem of ordered tests.³⁴⁰ Specifically, Dorf posits a case in which an appellate court determines that a lower court committed harmless error.³⁴¹ Because the finding of harmlessness results in an affirmance, the question is whether the finding of error, which but for the harmlessness would produce a remand and perhaps a reversal, is nonetheless a holding. Dorf concludes that it is because “[i]t forms an essential ingredient in the process by which the court decides the case, even if, viewed from a post hoc perspective, it is not essential to the result.”³⁴² Again, we agree with Dorf’s ultimate conclusion, but we find his analysis question-begging in much the same way that Dorf himself claims at a different point in his analysis that essentiality proves question-begging in the *Restatement*.³⁴³ If one determines that finding error is essential, that follows from the intuition that even if courts are permitted to assume error and then assess harmlessness, harmless error analysis in criminal appeals remains substantially ordered.³⁴⁴ Our analysis, in contrast, allows jurists and legal scholars to pinpoint exactly why the harmless error question is difficult, namely that one might dispute just how ordered the test is.

B. Necessary Conditions for a Holding

Having now considered several definitions of holding and dicta, we will set out our own. This definition seeks to capture in an easily applied manner the resolutions of the problem categories that we have suggested above.³⁴⁵ A holding consists of those propositions along the chosen decisional path or paths of reasoning that (1) are actually decided, (2) are based upon the facts of the case, and (3) lead to the judgment. If not a holding, a proposition stated in a case counts as dicta.

Each of the three numbered prongs is premised upon the power of the deciding jurist or court to select among one or more potential decisional paths in resolving a case. The first prong thus requires that for a proposition to be credited as a holding, it must be actually decided along

³⁴⁰ See *supra* Part II.A.2.b.i.

³⁴¹ “The appeals court rules: (1) the trial court erred by admitting the confession because it was coerced, but (2) the error was harmless beyond a reasonable doubt, so the conviction is affirmed. Is part (1) dictum?” Dorf, *supra* note x, at 2045.

³⁴² *Id.* The purpose of this example apparently is to reinforce Dorf’s conclusion that alternative rationales are not unique by illustrating an additional situation in which part of a court’s rationale is not essential to the result.

³⁴³ *Id.* at 2041.

³⁴⁴ The point is quite similar to that raised in the context of the strict scrutiny test. The word “harmless” modifies the word “error,” and so lawyers might conclude, rightly or wrongly, that it makes logical sense to see whether an error exists before determining whether it is harmless. See *supra* text preceding and accompanying note 74.

³⁴⁵ See *supra* Part III.C.

the chosen decisional path or paths.³⁴⁶ The second proposition then limits the permissible range of articulated propositions that can be credited as holding by insisting that they be grounded in the facts of the case. And the third proposition, by linking the ultimate disposition to the selected decisional path, both constrains and broadens judicial discretion. While this prong ensures that holdings are limited to issues that lead to the judgment, it also recognizes that particular findings might move us away from the judgment as the means of traveling along a path that eventually gets to that judgment. By analogy, a trip from a southern to a northern destination will sometimes require turns south, east, or west, depending upon the configuration of, and the relationships among, the various roads along the chosen route. So it is with legal reasoning. Thus, our definition allows a judge to draw one or more conceptual lines (or paths) from case facts to judgment, and to credit all issues resolved along the chosen path or paths to produce holdings. At the same time, it discounts as dicta issues on paths that do not originate in the facts in the case or lead nowhere.

Our construction of a definition around the concept of a decisional path reflects the insight that judges, while bound by the presentation of cases as the means for creating law, often retain substantial choice in the means of using cases to make law. Our analysis is consistent with the premises that courts and legislatures have distinct lawmaking functions and that judicial decision making is legitimated by the passive quality of resolving cases presented by actual litigants.³⁴⁷ Nonetheless, when a case admits of more than one path to a particular resolution, or a broader or narrow arc in forming the path from facts to judgment, we believe that judges should be afforded appropriate flexibility in crafting holdings when selecting the governing path or paths. Without this concession, we would be forced back to one of the definitions of the holding-dicta distinction that we have criticized above both on theoretical and practical grounds.³⁴⁸ If

³⁴⁶ We recognize that in selecting as part of our formulation the words “along the chosen decisional path or paths of reasoning” we are employing language that does not form part of commonplace definitions of holding. That said, we are confident that our chosen language better captures the intuition that underlies most judicial constructions of holding and dicta. In addition, while the term “path” is perhaps new to the problem of construing holdings, the concept is quite familiar to lawyers. For example, it is common to speak of a chain of reasoning, which like a path or paths of reasoning implies a progression of steps from facts to disposition. We have selected the term “path” rather than “chain” because a chain connotes links, and specifically, the risk that one link might be broken, thus affecting the disposition. Path instead emphasizes the choice among means of getting from one place (the case facts) to another place (the judgment). Because the determination that an assertion is a holding or is dicta does not turn on the power of the reasoning with respect to each step (an indefensible proposition in a poorly reasoned case can nonetheless be a holding), a weak link will not undermine the judge’s selection of the route from facts to disposition for purposes of identifying what constitutes a holding in a case.

³⁴⁷ See *supra* Part III.B.1.

³⁴⁸ See *supra* Part IV.A.

judges cannot choose among paths, there will be virtually no holdings in cases of even modest complexity, and if judges cannot choose the appropriate level of breadth, then cases will almost always be distinguishable based upon factual differences that most would agree have little or no relevance in terms of providing a material basis for different legal treatment.

While our definition is easy to articulate, each element proves more complex than it may first appear. In addition, while we are confident in the general structure and meaning of our proffered definition, we recognize that our approach does not and cannot be used to dictate results in every conceivable case. The inherent breadth of the terms that we employ will necessarily invite some good faith disagreements among courts in determining how to treat statements presented as holdings in a prior case. We do not consider this a weakness, because we are as concerned with facilitating meaningful analysis of dicta and holding, using a common framework, as we are in applying that framework to resolve all cases. Indeed, we have intentionally phrased our definition in sufficiently broad terms to permit disagreement on the closest analytical issues that we have identified. That said, we will offer proposed resolutions to concrete questions that our definition poses.

Finally, we note that as with other matters of judicial administration,³⁴⁹ institution-specific concerns might also affect the manner in which some courts treat dicta and holding. While we believe that our approach can be applied generally, we recognize that for some courts, some adaptation might be required. An important example is the United States Supreme Court. Arguably, because the Supreme Court has the power to define the rules governing the distinction between holding and dicta, when the Court explicitly provides that a statement is in one category or the other, the statement is in that category, regardless of whether the Court's conclusions accord with the general definitions of holding and dicta that we, or others, have offered. It might even be appropriate for the Court to stretch the definition of holding given the Court's unique role in judicial administration. Because the Court sits at the apex of numerous pyramidically structured judiciaries and must make and clarify law through a relatively small number of cases, the Court might require greater latitude than other courts in determining the scope of its

³⁴⁹ For example, justiciability doctrines in the federal courts are often said to emanate from the requirements of Article III. *See, e.g.,* Robert J. Pushaw, Jr., *Article III's Case/Controversy Distinction and the Dual Functions of Federal Courts*, 69 NOTRE DAME L. REV. 447 (1994) (considering justiciability doctrine specifically for the federal courts). We thus agree with Dorf, *supra* note x, that Article III similarly may have implications for the appropriate distinction between holding and dicta.

holdings.³⁵⁰ This might provide a normative justification, for example, for allowing Powell broader flexibility to define the holding in *Bakke* than would have been the case applying our general definition. We have focused on the application of the normative factors that we have identified to courts in general, rather than to particular courts, and we leave to future analysis whether these factors have different implications for particular courts.

1. *Actually Decided*

We turn now to the question of what it means for an issue to be actually decided. Within the context of a chosen decisional path, the requirement that a proposition be actually decided includes within a case's holding conclusions that courts explicitly reach. When an issue is actually decided along a chosen decisional path, the possibility of an alternative path does not undermine the status of those issues actually decided along that chosen path. Unlike the reconciliation approach, our definition credits as holding conclusions that a court does in fact reach, even if the court might have resolved a case by taking fewer paths or by choosing a narrower ground. Thus, alternative possible justifications, alternative justifications, and breadth do not provide bases for classifying issues actually decided as dicta.

As long as a judicial opinion reports a summary of facts and the disposition, the opinion provides at a least a facts-plus-outcome or reconciliation-based holding.³⁵¹ Even in the absence of any reasoning on a particular issue, a case at least stands for the proposition that under certain facts, a particular party should prevail on the relevant source of law upon which he or she relies. We recognize that there may be arguments favoring a limit on the precedential weight of cases

³⁵⁰ Michael Solimine has noted that the size of the Supreme Court's docket could affect the Court's ability to monitor inferior courts' compliance with Supreme Court doctrine. *See, e.g.*, Michael E. Solimine, *Supreme Court Monitoring of State Courts in the Twenty-First Century*, 35 IND. L. REV. 335, 358-59 (2002) (noting, however, that the Court has managed to control the state courts by considering only a very small percentage of state cases). If the Court encountered difficulty controlling the lower courts, whether because of increasing caseloads or increasing defiance, there might be an argument for a broader definition of "holding" for the Supreme Court. Alternatively, however, the Court's need to monitor courts while maintaining a small docket might call for broad deference to Supreme Court dicta. *See infra* note 414.

³⁵¹ A case might not qualify if it fails to meet the minimum conditions of internal consistency. Suppose, for example, the court acknowledges a general rule that on the facts of a case, a particular result should obtain, but then reaches another result without providing any justification, such as a factual or legal point, for that result. In this case, a precedent-setting court faces a dilemma: Is the rule announced by the court the holding, or is the inverse of the rule the holding since that is what the court did? In such a case, there is a strong reason to conclude that because the reasoning and outcome of the case are in direct opposition, the case does not clearly produce any holding at all. This is a very rare exception, however. It must be distinguished from the much more common situation in which a court produces reasons that the court thinks support the outcome it reaches, and a later court concludes that those reasons are bad or even point in the opposite direction properly conceived. Precedent means that even bad reasoning may control, so it is only in cases in which the reasoning cannot even be understood that the case fails to produce a holding.

that fail to articulate at least somewhat generalizable propositions. First, such cases may reflect less careful judicial attention than cases directly addressing propositions of law with clear applicability to multiple cases.³⁵² Perhaps courts tend toward case-by-case decision making that fails to result in more generalizable precedents when parties' briefs do not explore the broader implications of a case, or when the facts are such that judges deem the case an inappropriate vehicle through which to consider them.³⁵³ Second, narrow fact-based decision making frees judges to resolve cases where the facts or equities on the one hand and the law on the other pull strongly in opposite directions or where the law is open and substantial equitable arguments can be advanced on both sides.³⁵⁴ Providing precedential effect to such decisions might lead to bad decisions in other cases. This argument reflects the familiar maxim that hard cases make bad law.³⁵⁵

While these factors counsel in favor of denying precedential effect to decisions with fact-based reasoning, other factors pull in the opposite direction. First, granting judges the power to resolve cases without creating any precedent at all threatens to undermine the rule of law, as judges may resolve cases based on factors that they would not want to admit serve as the basis of their decisions. Second, narrow decisions are often more carefully considered and will produce only narrow precedents, thus leaving considerable leverage for future decision makers.³⁵⁶ When a court fails to offer any broad propositions for its holding, then a later court can distinguish that decision on the basis of any difference manifested by the facts. Although precedents will necessarily constrain deciding courts, facts-plus-outcome holdings leave courts with a great deal of flexibility.³⁵⁷ In our view, legal policy is best advanced by requiring judges, at least in cases with precedential value,³⁵⁸ to set precedents, but to give judges flexibility to offer narrow and easily distinguished precedents.

³⁵² Such an argument might help to explain the increasing use of unpublished slip opinions in the United States Courts of Appeals. See COHEN, *supra* note x, at 72-81 (documenting increased reliance upon slip opinions in Courts of Appeals).

³⁵³ Within the Supreme Court, the combined features of docket control and the ability to dismiss certiorari as improvidently granted (DIGs) reduce this risk relative to courts generally.

³⁵⁴ For an illustration, see *infra* Part V.C (discussing *Penn Central*).

³⁵⁵ See *Winterbottom v. Wright*, 10 W & M 109, 152 Eng. Rep. 402, 406 (Ex. of Pleas 1842) (Rolfe, B.) (“Hard cases, it is frequently observed, are apt to introduce bad law.”). Alternatively, this argument might flip the maxim to suggest the self-evident proposition that bad law makes cases hard.

³⁵⁶ See *supra* Part III.B.2 (discussing the case for judicial minimalism).

³⁵⁷ See *supra* Part IV.A.1 (arguing that a requirement of reconciling cases is a regime in which generalizations in judicial opinions never count as holdings).

³⁵⁸ We do not here enter the debate on whether judges should be permitted to label some opinions as not having precedential

While our definition allows a court to articulate broad propositions of law or make narrow holdings with fact-intensive reasoning, we do not suggest that determining which type of reasoning a court embraced will always be a straightforward task. Judges frequently debate the scope of past holdings, and whether a case stands for a broad or narrow proposition sometimes becomes clear only over time. Of course, part of the reason for such debate is that the line between holding and dicta is uncertain, and the purpose of this project is to provide more definitive guidance on that question. But it would be impossible to produce mechanical rules that would allow for unambiguous classification of every ambiguous judicial statement. Once again, our project is to classify statements *given* the resolution of ambiguities. Our project demands not just the prior resolution of semantic and syntactic ambiguities, but also the resolution of ambiguities about whether a court actually decided a particular issue, or instead resolved the case on some other ground.

There may also be ambiguities about when an issue that is *not* discussed is actually decided. Consider, for example, a case in which a court resolves the merits without even mentioning the issue of standing, despite the court's obligation to consider *sua sponte* issues of subject matter jurisdiction.³⁵⁹ The issue is close,³⁶⁰ but we do not think that such an opinion has actually decided the issue of standing. Although an opinion produces at least one holding on every issue considered, the court in this example has failed to consider an applicable issue, and so the only holding in such a case would be the holding on the merits. One might argue that such a case supports an argument that the plaintiff had standing. The court at least has revealed that it took for granted that standing existed, and this may make a later argument that would deny standing seem inconsistent with existing intuitions and understandings. But it would be a stretch to say that the court held that the plaintiff had standing, and in the Supreme Court, there is

value. *See, e.g.*, The Hon. Boyce F. Martin, Jr., *In Defense of Unpublished Opinions*, 60 OHIO ST. L.J. 177 (1999) (listing and responding to various criticisms of unpublished opinions).

³⁵⁹ *See, e.g.*, *Capron v. Van Noorden*, 6 U.S. (2 Cranch) 126 (1804) (providing a classic statement of this principle).

³⁶⁰ A still closer case would be one in which a three-part test is applicable but a court only discusses two of the three parts, even though all three logically were necessary to the disposition the court reached. In such a case, it may be that the court has thought the issue so straightforward as not to require any discussion, and there is thus an argument that the court has actually decided the issue. But it also may be that the court simply has forgotten one of the parts of the test. If that is so, then the court has not actually decided the omitted part. In practice, however, the court will at least have created a facts-plus-outcome holding, and thus the case at least implies the resolution of the omitted part, so the distinction may be of little practical consequence.

precedent establishing that only decisions that actually address and decide threshold issues create precedent with respect to those issues.³⁶¹

“Actually decided” does not mean “expressly stated.” The “actually decided” requirement is broader in the sense that it includes implicit holdings that a court never quite explicitly announces. But it is narrower in that it excludes some statements that figure in the reasoning of the case but that a fair construction of the opinion would not find to lie on the path from case facts to case disposition. An “actually stated” rule might prove overinclusive—and perhaps destabilizing—if construed to turn even high-level abstract reasoning into holdings. The “actually decided” approach permits a court to engage in abstract reasoning, perhaps even invoking abstract principles of justice,³⁶² without threatening to enshrine every proposition in that reasoning into law. A judge who claims analytical consistency with, or inspiration from, the likes of Aristotle or Kant, or who relies for insight upon such thinkers as Rawls or Dworkin, has not thereby elevated that philosopher’s observations to the status of law, although any concrete applications of theory to issues presented can rise to the status of holding provided that they otherwise meet our definition.

In addition, a judge’s selection of a particular interpretive methodology will not necessarily credit that methodological choice as a holding. For example, imagine a Supreme Court statutory interpretation case in which the language of the statute points in one direction, while the legislative history points in the opposite direction.³⁶³ If a majority of five looked to legislative history in construing the statute, while the remaining four justices in dissent eschewed reliance on legislative history, does this establish a precedent that judges must consider legislative history? We do not think so. We do think, however, that the case does at least imply that where text and legislative point in opposite directions, a court is permitted to consider both

³⁶¹ See, e.g., *Lewis v. Casey*, 518 U.S. 343, 352 n.2 (1996) (“[W]e have repeatedly held that the existence of unaddressed jurisdictional defects has no precedential effect.”) (citations omitted).

³⁶² Most cases invoking principles of justice do so only loosely, or borrow a concept without considering its full context. See, e.g., *Goetz v. Crosson*, 967 F.2d 29 (2d Cir. 1992) (citing JOHN RAWLS, *A THEORY OF JUSTICE* (1971), for the principle of a veil of ignorance, without considering the substance of Rawls’s argument in any more detail).

³⁶³ For an issue that may soon reach the Supreme Court and in which there is wide agreement that text and legislative history point in opposite directions, see *Olden v. Lafarge Corp.*, 2004 Fed. App. 0296P (6th Cir. 2004). *Olden* holds that the federal supplemental jurisdiction statute overrules a prior Supreme Court case, despite legislative history indicating that the drafters did not intend to do this. The Third Circuit also found the text and legislative history to point in opposite directions, but sided with legislative history over text. See *Meritcare Inc. v. St. Paul Mercury Ins. Co.*, 166 F.3d 214 (3d Cir. 1999).

sources and sometimes may choose the interpretation suggested by the legislative history over that set out in the text.³⁶⁴

One might label this more modest conclusion an “implied holding.” The question then arises how one determines the scope of a holding that is implied rather than expressly stated. Our view is that when a court selects a particular path from facts to judgment and the resolution of one step along that path is implied, the proper rule of construction for lower courts is to find, at a minimum that the court implicitly held on the narrowest proposition needed to render the opinion internally consistent. While a lower court could construe the implied holding, within the context of the overall opinion, to be broader, it would be improper for the court to construe the implied holding more narrowly or in opposite fashion. In addition, because the holding is implied, we do not believe that it has the same horizontal stare decisis effect on the issuing court as does an express holding. The issuing court is free to reconsider the impliedly resolved issue unconstrained by the presumption created in the implied holding in a future case that squarely presents the issue, even though the implied holding binds lower courts.³⁶⁵

By limiting the scope of implied holdings in this manner, our definition properly emphasizes the importance of providing judges more flexibility with respect to those holdings that are actually decided. The “actually decided” approach thus helps to make sense of the intuition that courts’ explicit statements about what cases hold retain special importance. These statements come in at least three forms. First, sometimes a court will label a proposition as a holding.³⁶⁶ We do not think that such a label can transform dicta into a holding, and our proposed approach provides a means of testing such assertions. When a statement offered as “holding” does not form an essential step on the path from operative facts to case disposition, then, consistent with the observed practices regarding the stare decisis norm,³⁶⁷ future courts may

³⁶⁴ An important implication of this analysis is that a definition offered by a court to distinguish holding from dicta might itself be dicta. See *United States v. Johnson*, 256 F.3d 893, 922 (9th Cir. 2001) (Gould, J., concurring) (arguing that comments by other judges on the distinction between holding and dicta themselves would not bind future panels assessing the holding-dicta distinction).

³⁶⁵ We draw a parallel with the narrowest grounds governing fractured panel decisions. See *Marks v. United States*, 430 U.S. 188 (1977). The *Marks* doctrine, which dictates that in a fractured panel case, the controlling opinion is the one consistent with the outcome that resolves the case on the narrowest grounds, binds lower courts, but does not prevent a future Supreme Court from reevaluating the issue absent the presumption of precedent, given that the holding was expressed in an opinion that lacked majority support. Similarly here, the implied nature of the holding is sufficient to ensure that lower courts credit at least the narrowest implied proposition as a holding, while allowing the issuing appellate court, which did not expressly resolve the issue, to reconsider it in a future case unconstrained by a presumption of precedent respecting that issue.

³⁶⁶ See, e.g., *infra* note 392 and accompanying text (providing an example).

³⁶⁷ See *supra* Part III.A.3 (providing a model showing that judges will have incentives to discount overly broad holdings).

disregard the labeling and treat the assertion as dicta. Assuming that the issue labeled “holding” does form an essential step on the selected path, however, then the “actually decided” approach allows the court to select the level of abstraction at which the issue is resolved. Again, consistent with observed stare decisis norms, this encourages careful articulation of reasoning, thereby promoting the important objective of doctrinal clarity. And of course our approach invites courts to use language, such as the magic words “we hold,” to make explicit that it has in fact resolved an issue. Ambiguity will thus arise only when a court has chosen, intentionally or inadvertently, not to do so.

Second, a court might explicitly disclaim a proposition as a holding. A common form of this is for a court to assume *arguendo* that a particular proposition holds and then resolve the case on that basis. For example, the Supreme Court might accept a case on only one issue and then simply assume without deciding that the lower court had made a proper decision on another issue.³⁶⁸ We cannot turn this assumption into a holding on the theory that the Court could have arrived at its result only by implicitly approving of the lower court holding. Courts generally construct assuming *arguendo* arguments precisely to demonstrate the nondispositive nature of the identified (and assumed) issue resolution. Sometimes, a court might disclaim any intent to cover a fact situation beyond that presently before it. For a famous illustration, consider the following assertion by the per curiam opinion in *Bush v. Gore*: “Our consideration is limited to the present circumstances, for the problem of equal protection in election processes generally presents many complexities.”³⁶⁹ Whether or not it was appropriate for the Court to do so,³⁷⁰ we believe that this conclusion established that the equal protection analysis produced no more than an easily distinguishable facts-plus-outcome holding.³⁷¹ Similarly, although he did not do so,

³⁶⁸ See, e.g., *Doe v. Chao*, 124 S. Ct. 1204, 1212 n.12 (2004) (“That issue is not before us, however, since the petition for certiorari did not raise it for our review. We assume without deciding that the Fourth Circuit was correct to hold that Doe’s complaints in this case did not rise to the level of alleging actual damages.”).

³⁶⁹ 531 U.S. 98, 109 (2000).

³⁷⁰ Compare ALAN DERSHOWITZ, *SUPREME INJUSTICE: HOW THE HIGH COURT HIBACKED ELECTION 2000* at 81-84 (2001) (arguing that the limitation was unprincipled), with Peter Berkowitz & Benjamin Wittes, *The Lawfulness of the Election Decision: A Reply to Professor Tribe*, 49 VILL. L. REV. 429, 448 n.72 (2004) (disagreeing with Dershowitz).

³⁷¹ See Vikram David Amar, *Adventures in Direct Democracy: The Top Ten Constitutional Lessons from the California Recall Experience*, 92 CAL. L. REV. 927, 956 n.129 (2004) (“[S]uch an admonition to a lower court should move the lower court judge to find the narrowest, most fact-specific, but still principled way to read the decision.”) (quoting *Bush v. Gore*, 531 U.S. at 109). In *Southwest Voter Registration Project v. Shelley*, 344 F.3d 882 (9th Cir. 2003), vacated by 344 F.3d 914 (9th Cir. 2003) (en banc), a Ninth Circuit panel did rely on *Bush v. Gore* to stay a special gubernatorial election in conjunction with an effort to recall the Governor. The Ninth Circuit, however, quickly took the case en banc and produced the opposite result. Notably, while the court cited *Bush v. Gore* in describing the panel decision, see *id.* 917, it found no equal protection without expressly distinguishing, or indeed even discussing, *Bush v. Gore*. See *Southwest Voter Registration Project*, 344 F.3d at 918.

Justice Powell in *Bakke* might have explicitly disclaimed his analysis of the Harvard plan as holding simply by indicating that he was not prejudging a future case challenging such a program, should one be adopted by a state institution of higher learning.³⁷²

Third, courts can make statements about the scope of previous courts' holdings. For example, a post-*Bakke* court could make an explicit statement that *Bakke* held that plus factor programs are narrowly tailored to advance the goal of diversity. A court might make this statement regardless of whether the post-*Bakke* court's analysis were consistent with the holding-dicta distinction that we advance here. Of course, to be persuasive, such a statement would require a foundation. A court also might label a proposition in a prior case dicta. These statements by the precedent-citing courts themselves might count as holding or dicta, depending on whether they independently satisfy the governing criteria that we set out in this section. But through this process, courts can over time reduce, even if they cannot eliminate altogether, ambiguities that the holding-dicta distinction may not entirely resolve.

2. *Based on Facts of the Case*

For a proposition to be a holding, it must be “based on facts of the case.” More formally, when a general proposition depends on a particular factual predicate, that factual predicate must be true on the facts of a case,³⁷³ or the factual predicate must be assumed to be true.³⁷⁴ This requirement should be one of our least controversial. In part, this is because it provides a straightforward way of formalizing *Black's Law Dictionary's* definition that “dictum” is a statement “by the way—that is, incidentally or collaterally, and not directly upon the question before the court.”³⁷⁵ As a positive matter, when a judicial statement transparently implicates facts

³⁷² Even with Justice Powell's failure to disclaim an intent to make a holding, his statements may have still been dicta. *See infra* Part V.A.

³⁷³ We do not intend to enforce a strict distinction between “facts” and “law” here, as sometimes the law may be a relevant fact in the sense that we are using it here. The Davis plan evaluated in *Bakke* might be seen as a fact, or because it was created by a governmental agency, a law. That should make no difference to our analysis.

³⁷⁴ In assessing a motion for summary judgment, for example, courts will consider the facts in the light most favorable to the nonmoving party. *See* FED. R. CIV. P. 56; *Celotex Corp. v. Catrett*, 477 U.S. 317 (1986) (elaborating the summary judgment standard). When a court subsequently grants summary judgment, the legal reasoning will depend on assumed facts. Courts, of course, commonly resolve legal issues by assuming the facts to be as one party argues, and that should be enough for the legal analysis to be considered based on the facts of the case.

³⁷⁵ BLACK'S LAW DICTIONARY 1102 (8th ed. 2004) (quoting WILLIAM M. LILE ET AL., BRIEF MAKING AND THE USE OF LAW BOOKS 304 (3d ed. 1914)).

not involved in the case, courts generally take any conclusions drawn from such discussions to be dicta.³⁷⁶

The most straightforward implication of this requirement is that a court cannot resolve hypothetical questions,³⁷⁷ or effectively do the same by making broad statements that arbitrarily string together separate categories of facts,³⁷⁸ and then claim holding status for that resolution. The *Bakke* Court, for example, could not have rejected the Davis plan on the ground that it was a quota and then also held that even a nonquota-based affirmative action program would be unconstitutional if employed by a public magnet high school. Because *Bakke* itself did not involve a public high school, and because it rested on the constitutionally problematic quota, such a statement would be dicta. A court could, however, base a holding on the absence of a fact, which of course is in itself a fact. For example, the *Bakke* Court could have held that if a quota-based affirmative action plan did *not* involve a public high school, then it was unconstitutional. Moreover, the *Bakke* Court could create a holding in ruling that because the Davis plan was *not* a plus-factor plan, it should count as an unconstitutional quota.

The “based on the facts of the case” requirement thus still affords a court substantial flexibility. The *Bakke* Court might reach the public high school case depending upon the chosen path and the level of abstraction for deciding the case. Had the Court said, for example, that equal protection prevents any use of race by a government official, and that for that reason the Davis program is unconstitutional, the now broader holding would presumably bring in our hypothetical magnet high school. While this suggests that the same result prohibited through forging distinct categories can be reached by expanding the level of breadth, there is no reason to fear a tendency toward case resolutions at the broadest levels. Had the *Bakke* Court resolved the case in the broadest possible manner, it might have prevented the later Supreme Court from issuing rulings allowing the benign use of race in government contracting based upon a lower threshold than strict scrutiny,³⁷⁹ at least without overruling this hypothetical *Bakke*. While *Adarand v. Peña* eventually rejected precedents allowing a threshold lower than strict scrutiny

³⁷⁶ See, e.g., *Gabbs Exploration Co. v. Udall*, 315 F.2d 37, 39 (D.C. Cir. 1963) (“[A]n expression as to the law based on other facts is regarded as dictum and not controlling on lower courts . . .”).

³⁷⁷ See *supra* Part III.C.5.

³⁷⁸ See *supra* Part III.C.7.

³⁷⁹ See *Metro Broadcasting, Inc. v. FCC*, 497 U.S. 547 (1990) (approving of minority ownership preference policy), *overruled in* *Adarand v. Peña*, 515 U.S. 200 (1995); *Fullilove v. Klutznick*, 448 U.S. 448 (1980) (finding constitutional minority business enterprise set-aside provision).

for challenges to benign federal race-based classifications,³⁸⁰ Justice Powell in *Bakke* was unwilling to close the door on all race-based policies.

3. *Leading to the Judgment*

The requirement that a statement lead along the decision path of reasoning to the disposition reinforces the approach that we have adopted for determining whether a statement is a presumptive holding, namely whether a statement creates a supportive proposition as we have defined that term.³⁸¹ We have combined the concept of a path or paths of reasoning with the phrase “leading to the judgment” to clarify that we mean to include as holdings some nonsupportive propositions that meet these criteria. As with a traveler who must initially go south on an eventual northern route, so too jurists frequently must establish nonsupportive propositions en route to their eventual judgment.

Thus, where a court undertakes a strongly ordered analysis and resolves one proposition in the opposite direction of the ultimate disposition, we find that proposition to lead, along the selected path, to the disposition. For example, where a court finds that a plaintiff has standing and then rejects the plaintiff’s claim on the merits, the standing inquiry leads to the disposition in the sense that it is a necessary first step to the subsequent inquiry.³⁸² This is so even if finding standing would not have been necessary had the court selected an alternative decisional path, including most obviously, basing the judgment directly on the absence of standing. As an additional example, where a court creates a test in a biconditional statement, and one half of that biconditional statement is a supportive proposition, we count the biconditional statement as a whole as leading to the disposition. We believe that combining the notion of a decisional path with the words “leading to” naturally embraces our earlier resolutions of these admittedly difficult issues, while still allowing some room for debate.

³⁸⁰ See *Adarand Constructors, Inc. v. Peña*, 515 U.S. 200 (1995).

³⁸¹ See *supra* II.A.2.a.4 (distinguishing supportive from nonsupportive propositions).

³⁸² At the same time, our definition excludes some presumptive holdings from holding status on the basis of the problem of ordered analysis. For example, where a court finds no standing and then rejects the claim on the merits, we would find the merits resolution not to lead to the disposition. See *supra* text accompanying notes 82-84 (describing the problem); notes 224-225 and accompanying text (offering a resolution of the problem).

V. APPLICATIONS

To demonstrate that our definition not only meets our normative concerns, but also functions well when applied to complex cases, we now revisit *Bakke*, and then introduce several additional cases. We have selected this group of cases both because they are familiar to many readers and because they embrace a range of analytical concerns that test the difficulties of any proffered definition of holding.

A. Bakke

We turn now to *Bakke*, the case with which we began our analysis and which we used to establish the various categories of judicial assertions that helped us explore which should be credited as holding and which should instead be discounted as dicta. Our definition makes the analysis of *Bakke* straightforward, though it does so in part by revealing that certain propositions in *Bakke* require subjective judgments before they can be classified as holding and dicta. Our goal is not so much to convince readers of our ultimate conclusion, as *Bakke* includes propositions that are very close to the conceptual line separating holding and dicta, but to identify the key issues and to offer insights that could be employed in a meaningful analysis of the most difficult questions that remain.

Continuing to rely upon Justice Powell's opinion as controlling,³⁸³ we conclude that the vast majority of those articulated propositions that we identified as potential holdings in *Bakke* do in fact meet our definition of holding. Consider, for example, Justice Powell's conclusion that remedying the present effects of past societal discrimination is not a compelling interest. Powell unmistakably actually decided the issue. The issue, moreover, was based on the facts of the case. Powell's analysis of the issue implicitly assumed, without deciding, that the program would remedy the present effects of past societal discrimination, and we have concluded that when a court assumes facts based on one party's argument, that suffices for the resulting legal argument to be considered based on the facts of the case.³⁸⁴ Finally, the resolution of the issue led to the disposition, as the issue reflected the court's choice of a path (focusing on the compelling interest prong rather than the narrow tailoring prong) to a result that was necessary for the disposition.

³⁸³ See *supra* notes 36-39 and accompanying text.

³⁸⁴ See *supra* note 374 and accompanying text.

Most of the other propositions in *Bakke* are clearly holding under similar analysis. The only propositions in *Bakke* that remain complicated involve Powell's discussion of the Harvard plan. Justice Powell's conclusion that diversity is a compelling interest counts as dicta under our definition, because it does not lead to the judgment. That conclusion would point toward the opposite disposition, and Powell's decision to resolve it thus does not reflect a decision merely to pick one or more paths from the case facts to the judgment. This conclusion, however, reflects our admittedly subjective judgment that strict scrutiny is not a sufficiently ordered test to make treatment of the compelling interest prong a holding when it points in the direction opposite Powell's conclusion. If one were to conclude that the test is ordered, then the identification of a compelling interest would logically precede the necessary narrow tailoring inquiry, and based upon that analysis would count as holding. Under our alternative analysis, treating the test as not strictly ordered, because Powell's broader conclusion that a plus-factor plan such as Harvard's would be constitutional followed from his compelling interest conclusion, that conclusion also does not lead to the disposition and thus also counts as dicta.

Powell's narrow tailoring analysis, however, should count as a holding. The issue appears to be actually decided, in that Powell has indicated precisely how a Harvard-like plan should be treated.³⁸⁵ We would reach a different result if Powell had included language suggesting that he was offering only a possible resolution of a hypothetical issue, or perhaps even if he had left ambiguous whether the issue was actually decided. Because Powell unmistakably resolved the issue and defended the resolution by explaining that the use of race in admissions was an issue in the court below, Powell in fact actually decided the issue. The propositions, moreover, were based on the facts of the case. We have noted that the absence of a fact itself counts as a fact, and Powell's reasoning was based on the implicit conclusion that the Davis plan was not a plus-factor plan. Finally, we have already concluded, although we concede that the issue is close, that the discussion of the Harvard plan effectively created a test for whether the use of race is narrowly tailored to the diversity objective,³⁸⁶ and because that test led to the disposition, it counts as a holding.

³⁸⁵ See *supra* note 53 and accompanying text (quoting the relevant passage of Justice Powell's opinion).

³⁸⁶ See *supra* Part III.B.6 (addressing the problem of biconditional statements).

In sum, we conclude that Powell’s conclusion that diversity is a compelling interest is dicta, but the conclusion that the test for determining narrow tailoring to further diversity based on whether a plan is a quota or a plus factor should count as a holding. Although we would not expect courts to employ the more complex system of classification set out in Part II, it is worth noting that the results here are consistent with those in Table 4, which summarizes that discussion. The analysis reveals that the problems of alternative possible justifications, alternative justifications, and of antecedent breadth are readily resolved applying our definition. The problem of ordered analysis turns out not to apply to the strict scrutiny test. The problem of biconditionals, meanwhile, means that the inverse statements of holdings count as holdings when the two are resolved together as a biconditional. And Justice Powell was entitled to issue a broad holding on narrow tailoring rather than a narrower one.

B. Roe

We will now analyze *Roe v. Wade*,³⁸⁷ based upon our definitions of holding and dicta.³⁸⁸ Although *Roe* presents some of the same issues as *Bakke*, it turns out to be simpler. The analysis reveals that there are no persuasive reasons to change any propositions from their initial presumptive status as holding or dicta. This should reinforce that *Bakke* presents an atypically high level of difficulty, and even cases with multiple-part holdings ordinarily can be easily sorted out.

In *Roe*, Jane Roe, a pseudonym for a pregnant woman who brought a class action lawsuit, was alone granted standing to challenge a Texas statute that criminalized procuring or attempting to procure an abortion, unless necessary to save the mother’s life.³⁸⁹ Writing for a majority of seven, Justice Blackmun not only struck down that statute,³⁹⁰ but proceeded to use the *Roe* case as a vehicle with which to articulate a comprehensive approach to assessing state abortion laws. In Section XI of the majority opinion, Justice Blackmun stated: “[t]o summarize and repeat:”, and then presented the synopsis of his detailed set of rules in outline form, including numbered

³⁸⁷ 410 U.S. 113 (1973).

³⁸⁸ To avoid unnecessary exposition in the text, we develop, primarily in footnotes, a more detailed analysis of *Roe* that is similar to our analysis of *Bakke* in Part II. This more detailed analysis verifies that our intuition captures the nuanced judgments that we have reached about individual categories of holdings.

³⁸⁹ 410 U.S. at 118-18 and 123-28.

³⁹⁰ *Id.* at 164 (“The statute, therefore, cannot survive the constitutional attack made upon it here.”).

parts and subparts.³⁹¹ Lest anyone doubt the Court's intent, the sentence following the outline referred to "[t]his holding."³⁹²

The outline itself spelled out *Roe*'s famous (or infamous) trimester framework. The Court began by identifying two legitimate state interests: protecting maternal health and the potential life represented by the fetus. The Court concluded that the first became compelling only after the first trimester, the point at which morbidity rates for carrying a fetus to term exceeded those for the termination of pregnancy through abortion. The Court further concluded that the second interest became compelling only at the moment of viability, the point at which the fetus is capable of meaningful life outside the mother's womb.³⁹³ After further concluding that the right to terminate a pregnancy falls within the right of privacy initially articulated in *Griswold v. Connecticut*,³⁹⁴ and extended to nonmarried couples through equal protection in *Eisenstadt v. Baird*,³⁹⁵ the Court weighed the two identified two state interests against the woman's right to abort. The Court then reached conclusions for each of the three identified trimesters of pregnancy.

In the first trimester, the right to abortion is absolute: "the abortion decision and its effectuation must be left to the medical judgment of the pregnant woman's attending physician."³⁹⁶ In the second trimester, "the State, in promoting its interest in the health of the mother, may, if it chooses, regulate the abortion procedure in ways that are reasonably related to maternal health."³⁹⁷ Finally, in the third trimester, "the State in promoting its interest in the potentiality of human life may, if it chooses, regulate, and even proscribe, abortion except where it is necessary, in appropriate medical judgment, for the preservation of the life or health of the mother."³⁹⁸

³⁹¹ *Id.* at 164-65.

³⁹² *Id.* at 165.

³⁹³ Justices have since pointed out that the point of viability may depend on developments in medical science. See *Akron Ctr. for Reprod. Health*, 462 U.S. 416, 458 (1983) (O'Connor, J., dissenting) (arguing that technology advances mean that "[t]he *Roe* framework... is clearly on a collision course with itself"). *But see Webster v. Reproductive Health Svcs.*, 492 U.S. 490, 554 n.9 (1989) (Blackmun, J., concurring in part and dissenting in part) (arguing that "the threshold of fetal viability is, and will remain, no different from what it was at the time *Roe* was decided").

³⁹⁴ 381 U.S. 479 (1965).

³⁹⁵ 405 U.S. 438 (1972).

³⁹⁶ *Roe*, 410 U.S. at 164.

³⁹⁷ *Id.*

³⁹⁸ *Id.* at 164-65.

The “actually decided” issue is easy to apply to the articulated *Roe* propositions because the Court made unmistakably clear that it intended its articulated rules for each of the three trimesters to count as holdings.³⁹⁹ Ambiguities might remain about other aspects of the opinion, however, such as the following statement: “This right of privacy, whether it be founded in the Fourteenth Amendment’s concept of personal liberty and restrictions upon state action, as we feel it is, or, as the District Court determined, in the Ninth Amendment’s reservation of rights to the people, is broad enough to encompass a woman’s decision whether or not to terminate her pregnancy.”⁴⁰⁰ The decision leaves some room for interpretation about whether the Court conclusively determined that the right of privacy is found in the Fourteenth Amendment. There is no doubt that the Court endorsed a view that the Fourteenth Amendment houses the privacy right—the Court’s language “as we feel it is” makes its view clear—but one might argue about whether the Court actually decided the issue for purposes of issuing a holding.⁴⁰¹ The Court might have been indicating a tentative view on the issue but not resolved it in the immediate case, perhaps based upon the intuition that recognizing a right—without regard to its doctrinal underpinnings—was adequate to the task of resolving the constitutionality of the challenged Texas abortion statute. Because the classification difficulty stems from the ambiguous wording of the opinion itself, we do not believe that our analysis resolves this specific question.

Roe might also present a problem under the “based on the facts of the case” prong. The Court resolved statements about each of the three trimesters of pregnancy, even though *Roe* herself must have filed suit during a particular single trimester. On this theory, the court might issue a single broad holding covering the first two trimesters, which together constitute the category of pre-viability, but cannot issue a separate holding for each of the three trimesters. We reject this analysis, however, and conclude that for two different reasons, all three trimesters were before the Court.

First, *Roe* was heard and decided after the plaintiff’s pregnancy had ended. Indeed, as a precondition to resolving *Roe* on the merits, the *Roe* Court articulated an exception to general mootness doctrine. Given the short duration of pregnancy, as compared with the protracted

³⁹⁹ See *supra* note 392 and accompanying text.

⁴⁰⁰ *Roe v. Wade*, 410 U.S. 113, 153 (1973).

⁴⁰¹ This would matter, for example, if a future case arose in which a privacy founded on the Fourteenth Amendment would afford protection, while a privacy right founded on another source would not.

process of litigation, the Court held that review is proper in an otherwise moot case when the issues are “capable of repetition, yet evading review.”⁴⁰² Accepting this justification for avoiding standard mootness doctrine, we can reconceptualize *Roe* as presenting a single plaintiff who, at progressive points in her pregnancy, is challenging the state law prohibitions against obtaining an abortion.⁴⁰³ Second, the Court found its reasoning sufficient to strike down the statute in its entirety.⁴⁰⁴ The Court thus implicitly found that Jane Roe presented a facial, rather than an as-applied, challenge to the statute. In a facial challenge, any constitutional defect is potentially sufficient to strike down the challenged statute, and, as a result, even hypothetical facts can be embraced within the facts of the case.⁴⁰⁵

Even though all three trimesters were properly before the Court, however, not all conclusions drawn concerning each trimester necessarily constitutes a holding under our definition. Under the third prong of our definition, to be a holding, a proposition must “lead to the judgment.” The *Roe* Court based its decision to struck down the Texas abortion statute on regulatory defects in the first two trimesters. The Court was certainly able to select more than a single path, provided each selected path led to the judgment. The difficulty is that Blackmun’s analysis of the third trimester abortion regulation did not lead to the judgment. The Texas statute did in fact contain an exception to the prohibition against abortion as needed to preserve the life or health of the mother. No path of reasoning that the court selected had as an essential element a finding that one aspect of the Texas abortion statute—in this case the exception for the life and health of the mother in the final trimester—withstood constitutional scrutiny.⁴⁰⁶

These conclusions withstand the type of closer analysis that we offered in *Bakke*. Rather than treat *Roe*’s articulated three-part holding on the constitutionality of abortion regulations at

⁴⁰² 410 U.S. at 125. This does not mean that anyone can raise an ephemeral issue in litigation. The issues must be capable of repetition yet evading review for the particular litigant who but for mootness satisfied all of the remaining justiciability requirements in the case.

⁴⁰³ The opinion itself notes that *Roe* was pregnant “as of the inception of her suit in March 1970 and as late as May 21 of that year.” 410 U.S. at 124. Because it does not indicate when she became pregnant, it is not clear from the facts of the opinion itself whether she was in the second trimester of her pregnancy at the time of the hearing.

⁴⁰⁴ 410 U.S. at 164 (“The statute, therefore, cannot survive the constitutional attack made upon it here.”).

⁴⁰⁵ Richard Fallon argues that whether a challenge is “facial” or “as applied” depends on the applicable constitutional doctrine, rather than on some trans-substantive rule. See Richard H. Fallon, Jr., *As-Applied and Facial Challenges and Third-Party Standing*, 113 HARV. L. REV. 1321 (2000). On this view, a court might strike down a statute based on hypothetical facts, but only because the plaintiff’s case presents facts that implicate doctrine that necessarily requires that such hypotheticals to be assessed.

⁴⁰⁶ Of course, the Court’s third trimester analysis might indicate that a hypothetical statute, one unlike the challenged Texas statute, would be unconstitutional. If so, the analysis fails under an alternative prong of our holding definition because it would not be based on the facts of the case. See *supra* Part IV.B.2. Under either analysis, the Court’s resolution of the third trimester counts as dicta.

various points in the pregnancy as a handful of holdings, one can break down that decision's nuanced test of constitutionality into several discrete propositions.⁴⁰⁷ This breakdown makes quickly apparent that the Court's conclusion that the right of privacy extends to abortion is a presumptive holding,⁴⁰⁸ as are almost all of the court's conclusions as to the first and second trimesters.⁴⁰⁹ In addition, the exceptions count as a holding based on our resolution of the problem of biconditional statements.⁴¹⁰ The conclusions relating to the third trimester, however,

⁴⁰⁷ The propositions are as follows:

- (1) Because the right of privacy extends to abortion, a state law restricting abortion must satisfy strict scrutiny to be sustained.
- (2) The state's interest in protecting maternal health is not compelling in the first trimester.
- (3) The state's interest in protecting maternal health is compelling after the first trimester.
- (4) The state's interest in protecting potential human life represented by the fetus is not compelling in the first and second trimesters.
- (5) The state's interest in protecting potential human life represented by the fetus is compelling in the third trimester.
- (6) Even when the state interest in the potentiality of human life represented by the fetus is compelling, the state may not pursue that interest at the expense of the life or health of the mother.
- (7) In the first trimester, the right to abortion is absolute.
- (8) In the second trimester, the state may regulate abortion in ways that are reasonably related to maternal health.
- (9) In the second trimester, the state may not regulate abortion in ways that are not reasonably related to maternal health.
- (10) In the third trimester, the state may regulate abortion through a statute that provides an exception to preserve the mother's life or health.
- (11) In the third trimester, the state may not regulate abortion through a statute that does not provide an exception to preserve the mother's life or health.
- (12) The statute regulates abortion in the first trimester.
- (13) The statute regulates abortion in the second trimester in a way that is not reasonably related to maternal health.
- (14) The statute regulates abortion in the third trimester, but provides an exception to preserve the mother's life or health.
- (15) The statute unconstitutionally regulates abortion in the first trimester.
- (16) The statute unconstitutionally regulates abortion in the second trimester.
- (17) The statute constitutionally regulates third-trimester abortion.
- (18) The statute is unconstitutional.

As before, we recognize that it would be possible to create a more detailed set of propositions, but we believe that we have included all the statements, some explicit and some implicit in the opinion, necessary to understanding the case. Note that the presentation above divides biconditional statements into two, so that we can evaluate each half of the biconditional statement separately. For example, statements (2) and (3) jointly form the biconditional "the interest in protecting maternal health is compelling if and only if the pregnancy is at a point after the first trimester," and so too with statements (4) and (5).

⁴⁰⁸ Statement (1) (right of privacy extends to abortion), is necessary to statements (7), (9), and (11), which identify particular statutory practices as unconstitutional.

⁴⁰⁹ Statement (2) (maternal health interest not compelling in first trimester) and statement (4) (interest in protecting fetal life not compelling in first and second trimesters) are necessary to statement (7) (right to first-trimester abortion is absolute). Statement (7) and statement (12) (statute regulates first-trimester abortion) are both necessary to statement (15) (statute unconstitutionally regulates first-trimester abortion), which is sufficient for statement (18) (statute is unconstitutional). Similarly, statement (4) (interest in protecting fetal life not compelling in first and second trimesters) is necessary to statement (9) (state may not regulate second-trimester abortion for reasons other than maternal health). Statement (9) and statement (13) (statute regulates abortion in the second trimester in a way that is not reasonably related to maternal health) are both necessary to statement (16) (unconstitutionally regulates second-trimester abortion), which is also sufficient for statement (18) (statute is unconstitutional).

⁴¹⁰ Statement (3) (state's interest in protecting maternal health is compelling after the first trimester) is necessary to statement (8) (state may regulate second-trimester abortion in ways reasonably related to maternal health). Statement (8) is a nonsupportive proposition, because it does not justify any other statement in the opinion or affect the need for any additional analysis, but it is the inverse of statement (9), and the two together create a test in the form of a biconditional statement. Statement (3) should thus count as a holding as well.

are nonsupportive propositions.⁴¹¹ Although one might argue that some of the key provisions can be rescued as inverse propositions to other holdings, we would disagree.⁴¹²

We do not mean to suggest that a post-*Roe* court confronted with a statute containing a third-trimester prohibition on abortion without an exception to preserve the life or health of the mother would have acted appropriately if it had glibly concluded that the Court's assertions were dicta and proceeded to ignore them. There are two reasons that a court arguably should not do this. First, even though our analysis suggests that the third-trimester propositions were not holdings, the Supreme Court's position at the apex of the judicial hierarchy may give it the power to define what is a holding and what is not.⁴¹³ Our analysis may suggest that the Supreme Court overreached, but a lower court ignoring the Supreme Court's delineation of holdings would be overreaching as well. Second, even Supreme Court dicta may have substantial persuasive influence on lower courts.⁴¹⁴ Once the Supreme Court has carefully considered and

⁴¹¹ Now, we can consider the propositions related to the third trimester. Statement (5) (state's interest in protecting fetal life is compelling after the second trimester) is necessary to Statement (10) (state may regulate third-trimester abortions if providing an exception to preserve mother's life or health). Statement (10) and statement (14) (statute regulates abortion in the third trimester while providing an exception to preserve mother's life or health) are necessary for statement (17) (the statute constitutionally regulates third-trimester abortion). Because the Court ultimately strikes down the statute, however, statement (17) is a nonsupportive proposition. Statement (3) (state's interest in protecting maternal health is compelling after the first trimester) and statement (6) (compelling interest in maternal health trumps compelling interest in fetal health) are both necessary to statement (11) (state may not regulate third-trimester abortions without providing an exception to preserve mother's life or health). Because the statute at issue does provide an exception, statement (11) is a nonsupportive proposition.

⁴¹² One might seek to label statement (10) (state may regulate third-trimester abortions if providing an exception to preserve mother's life or health) as an inverse proposition to a holding, and then statement (11) (state may not regulate third-trimester abortions if not providing an exception to preserve mother's life or health) as an inverse proposition to a holding. This analysis is somewhat more complicated, however. The Court effectively endorsed the proposition "if a pregnancy is in the first or second trimesters, then a ban on properly regulated abortions is unconstitutional." This entails the narrower proposition "if a pregnancy is in the first or second trimesters, then a ban on properly regulated abortions not needed to protect the mother's health or life is unconstitutional." The inverse of this proposition is "if a pregnancy is in the third trimester, then a ban on properly regulated abortions not needed to protect the mother's health or life is constitutional," which implies statement (10). By rewording the conditional in statement (10), we reach the proposition "if the state provides an exception to preserve the mother's life or health, then it may regulate third-trimester abortion." Note that the inverse of that statement—"if the state does not provide an exception to preserve the mother's life or health, then it may not regulate third-trimester abortion" is statement (11). If statement (10) indeed were a holding, we would agree that statement (11) would be a holding as well, because the Court has effectively announced a test for the third trimester. But because statement (10) is not part of any biconditional test that the Court has effectively announced—we can see that it is the inverse of a holding only by noting that the Court has endorsed a holding that entails a narrower proposition—we would conclude that neither it nor statement (11) should count as a holding.

⁴¹³ See *supra* note 350 and accompanying text.

⁴¹⁴ Commentators frequently stress the need for lower courts to give substantial deference even to Supreme Court dicta. See, e.g., *In re McDonald*, 205 F.3d 606, 612 (3d Cir. 2000) ("[W]e should not idly ignore considered statements the Supreme Court makes in dicta. The Supreme Court uses dicta to help control and influence the many issues it cannot decide because of its limited docket."). Consider also the following admonition:

A federal Court of Appeals is not bound by dicta in United States Supreme Court opinions, although such expressions are worthy of serious consideration. Federal appellate courts are, however, bound by the Supreme Court's considered dicta, almost as firmly as by the court's outright holdings, particularly when a dictum is of recent vintage and not enfeebled by any subsequent statement.

AM. JUR. 2D Appellate Review § 603 (2004).

resolved an issue, there are strong prudential reasons for lower courts to follow the Supreme Court's lead, even when that resolution technically would count as dicta.

C. Penn Central

The definition of holding or dicta is tested when applied to a case in which a court identifies a set of controlling factors, some pulling in one direction and others pulling in the opposite direction, and then balances these factors to rule for one side. While this can occur when legal doctrine points in one direction while the equities point in the other, it is most likely to occur when doctrine fails to provide clear guidance in the immediate case and when substantial equitable arguments can be advanced to support either side. The landmark takings case, *Penn Central Transportation Co. v. New York*,⁴¹⁵ illustrates the limited nature of holdings that arise in such balancing test cases.

In *Penn Central*, the Supreme Court addressed whether applying New York City's Landmark Preservation Law⁴¹⁶ to designate Grand Central Terminal an historical landmark and to prevent the construction of a high rise office building above the Terminal constituted a "taking" for which just compensation was required.⁴¹⁷ Justice Brennan, writing for a majority, divided the case into two issues.⁴¹⁸ Brennan first considered whether the designation constituted a taking for which just compensation was required because the burdens imposed upon Penn Central were greater than the generally shared burdens associated with zoning restrictions.⁴¹⁹ Concluding that the differential treatment did not automatically trigger a taking, Brennan then considered whether the particular burdens were sufficiently substantial to require an "an exercise of eminent domain and compensation to sustain [it]."⁴²⁰ The Court resolved both questions in the negative and thus found that there was no taking.⁴²¹

⁴¹⁵ 438 U.S. 104 (1978).

⁴¹⁶ N.Y. City Admin. Code ch. 8-A, § 205-1.0 et seq. (1976).

⁴¹⁷ See *Penn Central*, 438 U.S. at 122.

⁴¹⁸ *Id.* at 136.

⁴¹⁹ *Id.* at 114.

⁴²⁰ *Id.*

⁴²¹ Had the Court instead determined that the application of the Landmark Preservation Law to Penn Central constituted a taking, it would then have had to resolve the separate question whether the conferral upon Penn Central of transferable development rights (TDRs) satisfied the requirement of just compensation. See *id.*

The landmark designation was part of a city-wide scheme, similar to laws in all fifty states, pursuant to which several hundred historic properties were designated as landmarks and thus restricted from development without approval of the Landmarks Preservation Commission (the “Commission”).⁴²² A landmarks designation imposed considerable burdens on the owner both concerning the maintenance of the property and the process for approving any proposed developments.⁴²³ When the Landmarks Commission rejected each of two proposed plans to construct a multistory high rise above the Penn Central Terminal,⁴²⁴ appellants, Penn Central, and its lease partner, UGP Properties, Inc. (“UGP”),⁴²⁵ filed a suit challenging the application of the landmarks law to it as a taking.⁴²⁶

Justice Brennan placed the landmarks law between two sets of takings principles. The Court has generally not required just compensation in cases presenting challenges to zoning laws because zoning generally imposes approximately equal burdens on landowners, thus ensuring that a small group of owners are not unduly burdened for a public benefit.⁴²⁷ Conversely, the Court has found that compensation is required when the nature of the governmental action is such that the resulting harm undermines the owners’ investment-backed expectations.⁴²⁸ Examples include physical invasions of property or regulations that effectively thwart the owner’s ability to use the property in the manner that was anticipated based upon expressly retained rights.⁴²⁹ While the case law is not entirely consistent, the Court has found takings where

⁴²² *See id.* at 111.

⁴²³ *Id.* at 111-12.

⁴²⁴ *Id.* at 117. Pursuant to the landmarks law, Penn Central and UGP applied for a “certificate of no effect on protected architectural features,” which required that they demonstrate that the proposed changes “will not change or affect any architectural feature of the landmark and will be in harmony therewith.” *Id.* at 112. In addition they applied for a “certificate of appropriateness,” which required that they prove, based upon “aesthetic, historical, and architectural values,” that the proposed alterations “would not unduly hinder the protection, enhancement, perpetuation, and use of the landmark.” *Id.* The Landmarks Commission denied both permits for each proposed plan. *Id.* at 117.

⁴²⁵ In January 1968, Penn Central and UGP, a wholly owned subsidiary of Union General Properties, Ltd., a United Kingdom corporation, entered into a 50 year renewable lease and sublease agreement, under which UGP would construct a multistory office building atop the terminal and would pay Penn Central \$1 million per year during construction and at least \$3 million per year thereafter. These rents were intended to offset losses in rentals from concessionaires who would be displaced by the new building. *See id.* at 116.

⁴²⁶ Penn Central did not seek judicial review of these denials because Grand Central Station “remained suitable for its past and future uses, and was not the subject of a contract of sale,” and thus, Justice Brennan explained that “no further administrative remedies available to appellants as to [these] plans.” *Id.* at 118.

⁴²⁷ *Id.* at 123-24 (internal quotations and citations omitted).

⁴²⁸ *Id.* at 124.

⁴²⁹ In *Pennsylvania Coal Co. v. Mahon*, 260 U.S. 393 (1922), for example, the Court found a taking where, following the sale of property that included the reservation of mining rights, the state enacted a law preventing coal mining that caused subsidence of adjoining property given that the law directly thwarted the owner’s investment-backed expectations. In contrast, the Court did not find a taking in *Miller v. Schoene*, 276 U.S. 272 (1928), where a group of homeowners were forced to cut down ornamental

the state has engaged in a physical invasion of property,⁴³⁰ and where the challenged regulation entirely thwarts the expected use of property based upon the reserved rights of the property owner.⁴³¹

After holding that just compensation was not required any time a landowner is restricted in the use of property “to a greater extent than provided for under applicable zoning laws,”⁴³² Justice Brennan addressed whether “severity of the impact of the law on appellants’ parcel” was of such a magnitude as to rise to the level of a taking, thus requiring just compensation. Brennan resolved the latter issue against Penn Central by applying a balancing test.⁴³³

Brennan reasoned that because the landmark designation did not limit the “the present uses of the Terminal,” it did not undermine “Penn Central’s primary expectation concerning the use of the parcel.”⁴³⁴ Brennan added that the designation not only permits a profit from the Terminal, but also a “reasonable return” on Penn Central’s investment.⁴³⁵ In addition, Brennan observed that whatever loss appellants suffered from their inability to construct a high rise above the Terminal, the conferral of transferable development rights (TDRs), which creates the possibility for similar development on adjoining properties, provided some offsetting value, even if that value would not rise to the level of just compensation in the event of a taking.⁴³⁶ After concluding that the application of New York City’s Landmarks Law did not effect a taking for which just compensation is required,⁴³⁷ Brennan added that “our holding today is on the present record, which in turn is based on Penn Central’s present ability to use the Terminal for its intended purposes and in a gainful fashion.”⁴³⁸

red cedars that produced cedar rust fatal to nearby apple trees.

⁴³⁰ Thus, in *United States v. Causby*, 328 U.S. 256 (1946), the Court found a taking where the government had allowed flights directly over claimant’s land in a manner that prevented its then present use as a chicken farm.

⁴³¹ *Id.*

⁴³² *Penn Central*, 438 U.S. at 114.

⁴³³ *Id.* at 136.

⁴³⁴ *Id.*

⁴³⁵ *Id.*

⁴³⁶ *See id.* Brennan noted that the TDRs were “transferable to at least eight parcels in the vicinity of the Terminal, one or two of which have been found suitable for the construction of new office buildings.” *Id.* Brennan thus maintained that the TDRs “undoubtedly mitigate whatever financial burdens the law has imposed on appellants and, for that reason, are to be taken into account in considering the impact of [the] regulation.” *Id.* at 137.

⁴³⁷ *Id.*

⁴³⁸ *Id.* at 138.

Applying our definition, Justice Brennan chose a decision path from facts to judgment that required rejecting either of the two bases for ruling for Appellants, first finding that any substantial differential treatment among property owners, relative to traditional zoning laws, automatically constitutes a taking, and second, finding that on the specific case facts, the overall burden was sufficiently severe as to constitute a taking. Because a contrary ruling on either of these bases would have produced a judgment for Penn Central, and because each basis was actually decided based upon the facts of the case, applying our definition, each of these two bases is a holding.

In applying the balancing test used to resolve the second issue, Brennan identified a number of factors that satisfied the Court that the restrictions imposed were substantially related to promoting the general welfare and permitted the reasonable beneficial use of the Terminal and other properties. Justice Brennan identified the following burdens that the designation imposed on Penn Central: (1) the obligation to maintain the Terminal consistently with the requirements of the landmarks law; (2) the requirement that any development be approved by the Commission; and (3) the rejection of two proposals to construct a multistory office building above the Terminal, which would have enhanced the Terminal's profitability. Against these burdens, Justice Brennan weighed the following factors, which counseled against finding that the historic landmark designation imposed a sufficiently serious burden as to be treated as a taking: (1) that the designation did not interfere with the Terminal's primary use; (2) that the restrictions preventing the construction of the two proposed developments above the Terminal did not prevent a reasonable return on the owner's investment; (3) that while the Commission rejected two specific high rise proposals as inconsistent with the historic designated landmark, nothing in their actions suggested that they would altogether prevent any development, and specifically a more modest development, in the airspace above the terminal; and (4) that while they might not constitute just compensation, the landmark designation carried with it valuable TDRs.

Because the Court did not specify the weights that it placed on any single factor identified in its analysis, it is possible that in a case with even a single different factor or with the same factors in a situation warranting different emphasis based upon the particular case facts, a future court might obtain a contrary result. As a result, articulating the second *Penn Central* holding requires qualification based upon both the identified factors and the apparent emphasis

that the Court attributed to those factors based upon the specific case facts. For example, it is possible that on otherwise indistinguishable facts, the Court would find a taking if the Commission had prevented a substantially more modest construction in the air space or if the TDRs were less plausibly useful in adjoining properties, thus conferring Appellants with minimal or no value.

We do not wish to suggest that any distinction would be sufficient for a contrary result. A case with otherwise identical facts, but a proposal to build only a slightly less ambitious highrise than those which the Commission had rejected in *Penn Central* would appear insufficient to warrant a contrary result. Of course, courts sometimes do devise disingenuous distinctions, and balancing test cases might simply make that task easier. To the extent that an announced balancing test is intended to have meaning, however, we argue that an implicit substantiality requirement arises in efforts to distinguish future cases based upon differences in the weights attached to those factors that a court has identified in applying that test. And certainly over a group of cases, courts will give those factors more meaning, thus providing greater guidance to future courts.

It is also important to consider whether identifying the direction in which any given factor pulls when applying a balancing test itself represents part of the irreducible core holding. If for example a case presented otherwise identical facts except that the restriction involved a more ambitious construction plan, it would be hard to reconcile a holding that the new case presented a taking with *Penn Central*, which found no taking based upon a less restrictive imposition on the landowner. The converse proposition, however, does not hold. In a case that is otherwise indistinguishable from *Penn Central*, but in which the owner of the designated property did not receive any TDRs (which favor the property owner), it would be possible to rule either way without running afoul of *Penn Central*. Finding no taking would simply mean that the TDRs were relevant but not controlling in *Penn Central*, while finding a taking would mean that *Penn Central* was sufficiently close that changing this one factor changed the outcome.

Thus, where a factor in the balancing test weights in favor of the result (the presence of TDRs weights in favor of the result of finding no taking), future courts can reconcile the same or an opposite outcome in a case in which that factor is removed. Conversely, where a factor moves against the result (the extent of the restriction on construction in the airspace weighs against

finding no taking), a future court could not reconcile the opposite holding in a case in which the weight of that factor is exacerbated. As a result, the direction and weight of those factors that operate against the judgment in a balancing test case might themselves constitute a form of holding.

One might argue that even our conclusion concerning the direction in which an identified factor in a balancing test points does not constitute a holding that binds future courts. Consider again a case in which a court is presented with otherwise indistinguishable facts from those in *Penn Central* except that in the new case, the Commission denied construction of a very modest office building within the air space above the Terminal. Implicit in Justice Brennan's *Penn Central* analysis is the assumption that denying larger construction is less likely to constitute a taking than denying smaller construction. This assumption is consistent with Justice Brennan's concern, reflected in the actions of the Commission, that a larger construction is more likely than a smaller construction to compromise the physical appearance of the designated historical landmark.

It is possible to imagine a future judge, however, who views the relevance of the relative size of the proposed construction in nearly opposite fashion. In a future case that instead involves a relatively modest construction, the future judge might instead take the view that the relative value added to the Terminal resulting from the modest construction is sufficiently small that preventing it does not rise to the level of a taking, whereas the denial of a very large construction would have prevented such a substantial value added to the Terminal that it would have been more likely to constitute a taking. Justice Brennan and the hypothetical future judge have evaluated the significance of the relative scope of the proposed construction in the air space above the Terminal along two different analytical dimensions, and on the facts of the two cases, this difference in analytical approach potentially leads to opposite outcomes. The question then arises whether *Penn Central* "holds" that the future judge must rely solely on the analytical dimension that Justice Brennan concluded was significant (or at least on a dimension operating in the same direction), in this case the relative physical impact on the historic property, rather than on a dimension operating in the opposite direction, in the hypothetical case the relative value of the denied improvement to the overall property.

Applying our definition, Justice Brennan's choice of which analytical dimension to attach to a factor in a balancing test case, or, at a minimum, the choice of direction of that dimension, is a holding because it was based upon the case facts and led to the judgment denying the claimed taking. Because Justice Brennan articulated a test, his conclusion concerning the direction of this one factor constitutes a holding even though we do not know whether the application of that factor was itself dispositive. Here the direction of the factor for Brennan was the same as the judgment, thus mitigating any concern that he selected to attach a direction to this factor based primarily on his preferred view of legal policy. Given Justice Brennan's application of the balancing test we therefore conclude that it is not possible to reconcile with the holding in *Penn Central* an analysis that would select a dimension for analysis that points in the direction of using an expansive proposal as evidence in favor of a taking and evidence of a modest proposal against a taking. Had Justice Brennan reached the same result but focused on the analytical dimension emphasizing the relative value of the denied improvements, then the choice of that dimension would lead against the judgment, and it would count as dicta.

D. Blakely

The holding-dicta distinction often arises when a court must consider two cases from a higher court that are in tension with each other.⁴³⁹ Consider, for example, a debate in the aftermath of *Blakely v. Washington*,⁴⁴⁰ the recent case in which the U.S. Supreme Court struck down the State of Washington's sentencing guidelines system. Immediately after the issuance of *Blakely*, commentators recognized that there was a strong argument that the United States Sentencing Guidelines also violated the Sixth Amendment under the Supreme Court's reasoning.⁴⁴¹ The question remained, however, whether lower courts should have considered themselves free to consider the issue. Lower courts are required to follow all Supreme Court

⁴³⁹ In addition, it often arises when a court must consider two of its own cases that are in tension with one another. For example, in a recent case concerning whether U.S. district courts had habeas jurisdiction over detainees in Guantanamo Bay, Cuba, the Justices debated whether the Court had previously implicitly overruled a prior case. See *Rasul v. Bush*, 124 S. Ct. 2686, 2693-95 (2004) (arguing that *Braden v. 30th Judicial Circuit Ct.*, 410 U.S. 484 (1973), implicitly overruled *Johnson v. Eistrager*, 339 U.S. 763 (1950)), with *id.* at 2702-05 (Scalia, J., dissenting) (disagreeing with the majority). Case *B* can be thought of as overruling case *A* if a holding of case *B* and a holding of case *A* are inconsistent. Explicit application of the holding-dicta distinction that we develop here could discipline analysis in cases like *Rasul*.

⁴⁴⁰ 124 S. Ct. 2531 (2004).

⁴⁴¹ See, e.g., Stephanos Bibas, *Blakely's Federal Aftermath*, 16 FED. SENTENCING REP. 331 (2004).

precedents, even when manifestly inconsistent with more recent Supreme Court doctrine, unless and until the Supreme Court explicitly overrules its earlier precedent.⁴⁴²

The government took the position that this principle meant that the lower courts should not reconsider the constitutionality of the Federal Guidelines in the wake of *Blakely*.⁴⁴³ There had been numerous Supreme Court cases applying the Sentencing Guidelines, and litigants in one such case, *Edwards v. United States*, even raised Sixth Amendment issues.⁴⁴⁴ The Supreme Court in *Edwards* did not mention the Sixth Amendment, but did state that “we need not, and we do not, consider the merits of petitioners’ statutory and constitutional claims.”⁴⁴⁵ In a Seventh Circuit case considering this argument, Judge Posner concluded that *Edwards* did not create a holding on the *Blakely* issue, noting, “[a]n assumption is not a holding.”⁴⁴⁶ As Judge Posner’s use of the word “holding” reflects, a lower court faced with an apparent tension in the jurisprudence of a higher court must examine the higher court’s case law to identify what holdings the higher court has reached. As long as an earlier case does not contain a holding inconsistent with a later case’s holding, the lower court faces no dilemma about which case to follow.

We agree not only with Judge Posner’s recognition that the case presented a question of what counts as a “holding,” but also with his conclusion that there was no holding in *Edwards* that *Blakely* can be seen even to have implicitly overruled. In dissent, Judge Easterbrook thought that *Edwards* did bind the Seventh Circuit to continue to follow the Guidelines. “The Court’s opinion in *Edwards* acknowledged that constitutional contentions had been advanced,” Easterbrook argued. “*Edwards* held that a judge nonetheless may ascertain (using the preponderance standard) the type and amount of drugs involved, and impose a sentence based on that conclusion, as long as the sentence does not exceed the statutory maximum.”⁴⁴⁷ Under our definition, however, a court must *actually resolve* an issue for the case to create a holding, and when a court explicitly says “we need not, and we do not, consider the merits of petitioners’ statutory and constitutional claims,” any subsequent explanation that the claims do not seem

⁴⁴² See *State Oil Co. v. Khan*, 522 U.S. 3, 20 (1997); *supra* note 112.

⁴⁴³ Memorandum to All Federal Prosecutors from James Comey, Deputy Att’y Gen., July 2, 2004, available at <http://talkleft.com/dojmemojuly2.pdf>.

⁴⁴⁴ *Edwards v. United States*, 523 U.S. 511 (1998); see also Brief for Petitioners at 30-32, 1997 WL 793079, *Edwards* (No. 96-8732).

⁴⁴⁵ 523 U.S. at 515.

⁴⁴⁶ *United States v. Booker*, 375 F.3d 508, 514 (7th Cir. 2004).

⁴⁴⁷ *Id.* at 516 (Easterbrook, J., dissenting).

strong must count as dicta. Under our analysis, then, lower courts remained free even before *Blakely* to rule that the Sentencing Guidelines were unconstitutional for Sixth Amendment reasons.

That does not necessarily, however, justify Judge Posner's decision to find the Sentencing Guidelines unconstitutional, even assuming that this is the logical import of the *Blakely* decision.⁴⁴⁸ There may be prudential reasons for a lower court to honor Supreme Court dicta,⁴⁴⁹ and there may be even stronger prudential reasons for a lower court to honor the taken-for-granted assumptions of higher courts. There were at least signals in Supreme Court cases that the Justices did not think the Sixth Amendment argument against the Sentencing Guidelines meritorious. On the other hand, federal courts are obligated to honor a convicted defendant's constitutional rights even if in doing so they effectively force the Supreme Court's hand in the resolution of an issue that it previously assumed had no merit. What matters for our purposes is that our definition provides a way of determining when a lower court has the power to follow the logical implications of recent Supreme Court decisions over seemingly inconsistent earlier decisions. That issue does not resolve when a lower court should exercise the power if it exists.

E. *Grutter* and *Gratz*

While the recent *Grutter*⁴⁵⁰ and *Gratz*,⁴⁵¹ each issued by majorities, effectively remove any doctrinal significance from the inquiry whether Justice Powell's independent analysis in *Bakke* was dicta or holding, the holding-dicta distinction remains important for affirmative action law. Consider, for example, Justice O'Connor's statement in *Grutter* that "[w]e expect that 25 years from now, the use of racial preferences will no longer be necessary to further the interest approved today."⁴⁵² Is this bold assertion holding or dicta? One problem in answering this question is that it is not at all obvious even what Justice O'Connor's assertion means. It might even have been purposely ambiguous.⁴⁵³ It might not be surprising therefore that Justices Ginsburg⁴⁵⁴ and Thomas⁴⁵⁵ disagreed as to O'Connor's meaning.

⁴⁴⁸ At least one circuit court reached the issue of the Sentencing Guidelines' constitutionality but found that they were constitutional despite *Blakely*. See *United States v. Piniero*, 377 F.3d 464, 465 (5th Cir. 2004).

⁴⁴⁹ See *supra* note 414.

⁴⁵⁰ 539 U.S. 306 (2003).

⁴⁵¹ 539 U.S. 244 (2003).

⁴⁵² 539 U.S. at 342.

⁴⁵³ Justice O'Connor may have wanted to send a signal stronger than a mere expression of hope, yet recognized that the Supreme

Our analysis of holding and dicta can, however, help determine the consequence of any suggested interpretation. If, for example, the sentence is interpreted to mean that affirmative action survives for twenty-five years *and no more*, then O'Connor might be attempting to issue a biconditional statement that transforms an "if, then" assertion into an assertion that "if and only if a Michigan Law School-like affirmative action program is in place *no later than 2028*, is it constitutional." This, however, would seem to be a biconditional statement that provides an end-run around the exclusion of hypotheticals from holding, so there is a strong argument that the no-later-than-2028 restriction is not based on the facts of the case and counts as dicta. If, meanwhile, O'Connor's statement is interpreted as a mere personal hope or prophesy, then the issue is not actually decided. What is a holding today should remain so in 2028, and the statement setting out a time limit itself should be relegated to the status of dicta. Such dicta, of course, may have inspirational but not precedential effect.

VI. CONCLUSION

Our own hope for the future is modest. In this Article, we have argued that the distinction between holding and dicta helps to prevent judges from evading the principle of *stare decisis*. If judges were uniformly to adopt our definition and our proposed resolution of particular problems, courts would apply the holding-dicta distinction more consistently and in a manner more faithful to the underlying function of the distinction. Uniform adoption is unlikely at best, however, and many courts will likely continue to cite the *Black's Law Dictionary* definition of the distinction. We recognize too that the more definitions that judges rely upon, the easier it will be for them to justify their treatment of disfavored precedent. We believe that our definition

Court, unlike Congress, does not provide bright-line sunset provisions. See Vikram David Omar & Evan Caminker, *Constitutional Sunsetting?: Justice O'Connor's Closing Comments in Grutter*, 30 HASTINGS L.Q. 541, 554 (2003) (asserting that "[w]hile Congress occasionally enacts sunset provisions into its work product, the Supreme Court usually does not.").

⁴⁵⁴ "From today's vantage point," Justice Ginsburg wrote, "one may hope, but not firmly forecast, that over the next generation's span, progress toward nondiscrimination and genuinely equal opportunity will make it safe to sunset affirmative action." 539 U.S. at 346 (Ginsburg, J., concurring). By casting the uncertainty as whether we can forecast the end of a need for affirmative action or merely hope for such an end, Ginsburg signals that she does not view O'Connor's remarks as providing for an end to affirmative action regardless of the developments in the interim.

⁴⁵⁵ Justice Thomas made his interpretation of Justice O'Connor's comment patent: "I agree that in 25 years the practices of the Law School will be illegal," adding, "No one can seriously contend, and the Court does not, that the racial gap in academic credentials will disappear in 25 years. Nor is the Court's holding that racial discrimination will be unconstitutional in 25 years made contingent on the gap closing in that time." 539 U.S. at 344 (Ginsburg, J., concurring). Justice Thomas also seizes on a quotation included in Justice O'Connor's opinion, that the "the acid test of [affirmative action preferences'] justification will be their efficacy in eliminating the need for any racial or ethnic preferences at all." *Id.* at 342-33 (Thomas, J., dissenting, citing Nathaniel L. Nathanson & Casimir J. Bartnik, *The Constitutionality of Preferential Treatment for Minority Applicants to Professional Schools*, 58 CHICAGO BAR REC. 282, 293 (May- June 1977)).

captures the essential intuitions necessary for a theoretically satisfying and functional understanding of holding and dicta, one that we believe has the potential to benefit the process of judicial decision making. Whether or not courts use our precise language, we hope that over time they incorporate the valuable intuitions that we have expressed in this article, thus moving toward a more clear and coherent understanding of the important distinction between holding and dicta.