COMMENTS

Necessity in Free Exercise

Brady Earley†

The Free Exercise Clause is a broadly worded constitutional prohibition against government intrusion on religious exercise. To construct limits, courts have consistently required government officials to demonstrate the necessity of state action burdening religion. Yet government officials regularly fail to produce evidence of necessity, leaving judges to intuit or assume whether necessity exists. This Comment offers a better way. Using a method known as difference-in-differences (DiD), lawmakers can draw upon the experience of existing state laws to enact laws justified with evidence. This Comment demonstrates the value of DiD with a current free exercise controversy involving the Old Order Amish and their objection to Ohio's flashing light requirement for buggies. Applying DiD to this conflict reveals that Ohio's buggy light law led to an estimated 23% reduction in buggy-related crashes compared to laws in Michigan and Kentucky—states with less restrictive buggy requirements. Beyond this case study, this Comment also discusses how DiD can help address recent Supreme Court conflicts over tax exemptions for religious organizations, LGBTQ-themed books in schools, and religious charter schools. These examples grapple with the problems and showcase the possibilities of a data-driven method to address necessity in free exercise.

† B.S., B.A. 2021, Brigham Young University; J.D., Ph.D. Candidate in Political Science, The University of Chicago. I would like to thank Professor Geoffrey Stone and members of *The University of Chicago Law Review* including Owen Hoepfner, Jack Brake, Hannah Zobair, Ryan Jain-Liu, Zoë Ewing, Jackson Cole, and others for contributing to the publication of this Comment.

INTRODUCTION

From the beginning, the U.S. Constitution has largely been a product of compromise; many details were intentionally left undecided at the Founding. This lack of detail is certainly present in the Free Exercise Clause.[1] The Clause itself took on an amalgamated meaning from existing state constitution free exercise clauses and passed through Congress with relatively little debate.[2] While scholars have debated the precise meaning of the words "Congress shall make no law . . . prohibiting the free exercise [of religion],"[3] the more pressing doctrinal issue is the standard to apply in free exercise disputes based on existing Supreme Court precedent. This issue is challenging given the ambiguity in a broadly worded provision like "prohibiting the free exercise" of religion. While almost all can agree that the Free Exercise Clause is not an absolute protection from government action, the boundaries of the Free Exercise Clause's protections remain contested.

In searching for this boundary, the Supreme Court's free exercise doctrine has boiled down to a key evidentiary question: Was the government action burdening religious exercise necessary? When the government can demonstrate how burdening religion is necessary to achieve a valid public aim, courts have generally

---

1 *See* STEVEN D. SMITH, THE RISE AND DECLINE OF AMERICAN RELIGIOUS FREEDOM 48–75 (2014) (arguing that compromise over the text of the First Amendment was an intentional move to allow for open debate about how the Clause should develop).

2 *See* Michael W. McConnell, *The Origins and Historical Understanding of Free Exercise of Religion*, 103 HARV. L. REV. 1409, 1456, 1480–85 (1990) [hereinafter McConnell, *The Origins and Historical Understanding*].

3 U.S. CONST. amend. I, cl. 1. *Compare* McConnell, *The Origins and Historical Understanding*, *supra* note 2, at 1461–66 (pointing to free exercise exceptions for "peace and safety" in early state constitutions), *with* Philip A. Hamburger, *A Constitutional Right of Religious Exemption: An Historical Perspective*, 60 GEO. WASH. L. REV. 915, 918–26 (1992) (arguing that such exceptions were far more permissive of state reasons to burden religion than McConnell suggests).

deemed such action constitutional.[4] State actors can usually point to some public interest served by their laws. Government officials have argued that prohibiting polygamy was necessary to upholding public morality[5] and that restricting religious gatherings was essential to public health in a pandemic.[6] But when are these arguments valid, and how will courts know?

A recent controversy in Ohio illustrates the challenge of finding these appropriate boundaries of the Free Exercise Clause's protections. In 2022, Ohio passed H.B. 30[7] requiring all horse-drawn buggies to display flashing lights.[8] Some Amish groups in Ohio object to using electricity on religious grounds and therefore challenged the constitutionality of the flashing light requirement. In response, Ohio argued that the Free Exercise Clause allows the state to enact this law because the law is necessary to protect public safety, even if it means burdening the Amish's religious exercise to achieve that goal. How do courts decide whether the law is actually necessary? Because this dilemma recurs across free exercise cases, it is critical for courts to find a principled way of constructing constitutional meaning in these decisions.

This Comment offers a workable approach to this problem that is in line with free exercise precedent and the judicial capacity for evaluating evidence. The core inquiry in free exercise cases is whether government action that burdens religion is necessary.[9] But government defendants often fail to bring evidence of the necessity of their actions.[10] When this occurs, judges are left to intuit necessity through history, precedent, or other means. Consequently, judges are more likely to find or assume convenient answers rather than correct ones. An empirical method known as difference-in-differences (DiD) would allow lawmakers to better assist judges in evaluating evidence of necessity. This method compares the efficacy of a challenged state action, relative to its stated goals, to the efficacy of less restrictive alternative actions.[11]

---

4 *See infra* Part I.

5 *See* Reynolds v. United States, 98 U.S. 145, 166–67 (1878) (finding a government ban on polygamy constitutional after it was challenged on free exercise grounds).

6 *See* Tandon v. Newsom, 141 S. Ct. 1294, 1297 (2021) (per curiam) (holding that plaintiffs were likely to succeed on their free exercise claims that certain restrictions on religious gatherings during the COVID-19 pandemic were unconstitutional).

7 OHIO REV. CODE ANN. § 4513.114 (West 2025).

8 *See infra* notes 138–45 and accompanying text.

9 *See infra* Part I.

10 *See infra* Part II.A.

11 *See infra* Part II.B.

In the context of the Amish flashing light law, DiD could compare the impact of flashing lights on buggy accidents in Ohio to the impact of looser regimes (e.g., only requiring reflective markers) on buggy accidents in other states.[12] The DiD method brings analytical rigor to the necessity analysis while being more accessible to litigants and judges than more sophisticated techniques such as randomized controlled trials and field experiments.[13] When burdening free exercise of religion is necessary to achieve governmental aims, less restrictive alternatives will lack the efficacy of the challenged policy. By comparing outcomes from differing policy regimes, courts and legislatures can leverage the federalist system to operationalize the necessity test and protect free exercise in accordance with their institutional roles.

This Comment proceeds as follows. Part I traces the necessity principle in free exercise jurisprudence through three phases of constitutional history: the quiet early years, the high-water mark of strict scrutiny, and the current neutrality and general applicability regime under *Employment Division v. Smith*.[14] Part II turns to operationalizing necessity in free exercise jurisprudence. It begins by raising the evidentiary problems that come from a necessity test and then discusses how a solution may come from the DiD method. Part III then uses DiD to evaluate the necessity of Ohio's flashing light requirement for traffic safety. The analysis finds that Ohio's law produced an estimated 23% decrease in buggy-related crashes relative to a less restrictive alternative—strong evidence that the necessity test is satisfied. Part IV then considers how DiD may be valuable to assessing necessity in three free exercise cases the Supreme Court recently considered. At a time of uncertainty in free exercise doctrine, this Comment urges courts to pay greater attention to free exercise cases by employing rigorous evidentiary methods consistent with the roles of judges and litigating parties.

I. CONSTITUTIONAL HISTORY OF NECESSITY IN FREE EXERCISE

Since the Founding, government action burdening religious exercise has required a justification to comport with the Free Exercise Clause. This Part sets forth how courts have relied on a necessity inquiry to determine whether such government action

---

12 *See infra* Part III.B.

13 *See infra* Part II.B.

14 494 U.S. 872 (1990).

is justified. Part I.A begins with the early years of the Founding through the Supreme Court's 1878 decision *Reynolds v. United States*,[15] when the necessity principle operated largely in the background. Next, Part I.B describes the cases that brought necessity to the forefront of free exercise jurisprudence: *Sherbert v. Verner*[16] and *Wisconsin v. Yoder*.[17] Finally, Part I.C articulates the shift in *Smith*, in which "neutral law[s] of general applicability" became a shortcut for necessity in free exercise.[18]

A. Necessity in the Background: From the Founding to *Reynolds*

The text of the U.S. Constitution's Free Exercise Clause is a categorical statement: "Congress shall make *no law* . . . prohibiting the free exercise [of religion]."[19] Despite this text, there is almost universal agreement that the right to free exercise of religion is not absolute: Congress and other government actors are sometimes permitted to prohibit free exercise.[20] In fact, early federal and state constitutional free exercise decisions allowed the government to do so if burdening religion was necessary for the public good.

This necessity test—a generalized precursor to the narrow tailoring prong of strict scrutiny—operated as a background principle in the first federal case inviting arguments on the Free Exercise Clause,[21] *Permoli v. Municipality No. 1*.[22] Catholic priests in New Orleans challenged a city law prohibiting the performance of Catholic last rites during a public health crisis.[23] New Orleans's defense relied on a necessity test, explaining that "all obituary rites and ceremonials which tend to frustrate [the City's]

---

15 98 U.S. 145 (1878).

16 374 U.S. 398 (1963).

17 406 U.S. 205 (1972).

18 *Smith*, 494 U.S. at 879 (quoting United States v. Lee, 455 U.S. 252, 263 n.3 (1982) (Stevens, J., concurring in the judgment)).

19 U.S. CONST. amend. I, cl. 1 (emphasis added).

20 *See Reynolds*, 98 U.S. at 166 ("Suppose one believed that human sacrifices were a necessary part of religious worship, would it be seriously contended that the civil government under which he lived could not interfere to prevent a sacrifice?").

21 *See* Michael W. McConnell, *Schism, Plague, and Last Rites in the French Quarter: The Strange Story Behind the Supreme Court's First Free Exercise Case*, *in* FIRST AMENDMENT STORIES 39, 40 (Richard W. Garnett & Andrew Koppelman eds., 2012) (denoting *Permoli v. Municipality No. 1*, 44 U.S. (3 How.) 589 (1845), as the "first case to reach the United States Supreme Court in which a party explicitly invoked the protections of the Free Exercise Clause of the First Amendment").

22 44 U.S. (3 How.) 589 (1845).

23 *Id.* at 590 (quoting Municipality No. 1, New Orleans, La., Ordinance (Oct. 31, 1842)).

objects, or impair its efficacy, must yield to the supremacy of the common good."[24] Thus, while the text of the Free Exercise Clause did not spell out this exception, the public health necessity of the law was the central argument for its constitutionality. The Supreme Court dismissed the case for jurisdictional reasons, leaving these merits arguments untouched for the time.[25] It was not until a century later that a Supreme Court Justice approvingly cited the necessity rationale advanced in *Permoli* to support the shift toward heightened scrutiny of state action infringing on religious exercise.[26]

Reviewing Founding Era state constitutional free exercise clauses and case law provides additional support for the understanding of the federal free exercise protections argued for in *Permoli*.[27] A common element of these state constitutional provisions was a textually defined limit on free exercise in the interest of "peace," "safety," or some other aspect of public welfare.[28] Even when state constitutions were not explicit about these limits, state courts regularly read in such limits,[29] implicitly embracing a necessity test for defining the boundaries of constitutional protections for religious freedom.

For example, a familiar type of case brought under state constitutional protections involved nonmajority faiths challenging Sunday closing laws. *Specht v. Commonwealth*[30] illustrates a familiar fact pattern for these cases. There, a Seventh-Day Baptist named Jacob Specht argued that Pennsylvania's Sunday closing laws unconstitutionally burdened his religious exercise. Because Specht's Sabbath fell on Saturday, he argued that penalizing him for opening his business on Sunday violated the Pennsylvania Constitution's prohibition against "interfer[ing] with the rights of conscience."[31] Unmoved, the Pennsylvania Supreme Court upheld the law. The court explained that the state constitutional protection for religious freedom only covered that "which is not prejudicial

---

[24] *Id.* at 601.

[25] *Id.* at 610.

[26] *See* Braunfeld v. Brown, 366 U.S. 599, 613 (1961) (Brennan, J., concurring and dissenting) (citing *Permoli*, 44 U.S. at 600; United States v. Carolene Prods. Co., 304 U.S. 144, 152 n.4 (1938)).

[27] *See* McConnell, *The Origins and Historical Understanding*, *supra* note 2, at 1455–56.

[28] *See id.* at 1455–56, 1461–62 (discussing, for example, the New York Constitution's limit on religious protections found "inconsistent with the peace or safety of [the] State").

[29] *See* Commonwealth v. Wolf, 3 Serg. & Rawle 48 (Pa. 1817) (interpreting PA. CONST. of 1790, art. IX, § 3); City of Cincinnati v. Rice, 15 Ohio 225 (1846) (interpreting OHIO CONST. of 1802, art. VIII, § 3).

[30] 8 Pa. 312 (1848).

[31] *Id.* at 314 (quotation marks omitted) (quoting PA. CONST. of 1838, art. IX, § 3).

to the public weal."[32] Finding that days of rest were "absolutely necessary" to the "well-being of society," the court concluded that the Sunday closing law was constitutional.[33] The court recognized that the law was enacted "[i]n a Christian community" where "a very large majority of the people celebrate the first day of the week as their chosen period of rest from labour."[34] Thus, because the law only accomplished its purpose of a "respite from labour at the same time" by appealing to the masses, the Sunday closing law was deemed constitutional.[35]

Other state courts came to similar conclusions, implicitly noting the necessity of various government acts. In *Gabel v. City of Houston*,[36] the Texas Supreme Court made an even stronger appeal to the necessity of Sunday closing laws. The Texas court argued that these laws, which closed drinking establishments on Sundays, prevented the dangerous scenario in which "crowds of persons may be congregated at a public house, and, under the influence of intoxication, may commit riots and breaches of the peace."[37] Likewise, the Supreme Court of Alabama, in upholding a closing law, declared that "the general welfare and the good of society require a suspension of labor and business for one day in seven, and that [ ] day should be one of *uniform observance*."[38] When state courts seemingly deferred to legislative judgment on these issues, that deferral was animated by a background assumption that such laws were necessary from the start.[39] While they could have simply upheld these laws as a suitable exercise of legislative action, courts took pains to emphasize such laws

---

[32] *Id.* at 322 (quoting Commonwealth v. Lesher, 17 Serg. & Rawle 155, 160 (Pa. 1828) (Gibson, C.J., dissenting)).

[33] *Id.* at 323.

[34] *Id.*; *see also Wolf*, 3 Serg. & Rawle at 51 (concluding that "[l]aws cannot be administered in any civilized government unless the people are taught to revere the sanctity of an oath, and look to a future state of rewards and punishments for the deeds of this life" and that ensuring a Sunday Sabbath was of the "utmost moment" for inculcating such values); *Rice*, 15 Ohio at 242 ("[A]lthough . . . we may shut up the shops of the Jewish merchants in Cincinnati, on the first day of the week, we shall open the shop-doors of the whole mercantile community, in every other part of the State.").

[35] *Specht*, 8 Pa. at 323; *see also Wolf*, 3 Serg. & Rawle at 51 (finding that the right of conscience does not protect those who "directly oppose those laws, for the pleasure of showing their contempt and abhorrence of the religious opinions of the *great mass* of the citizens" (emphasis added)).

[36] 29 Tex. 335 (1867).

[37] *Id.* at 347.

[38] Frolickstein v. Mayor of Mobile, 40 Ala. 725, 728 (1867) (emphasis added).

[39] Importantly, that background assumption also included a "vast majority of [ ] people . . . belie[ving] in the Christian religion." *Gabel*, 29 Tex. at 345.

were not just appropriate but necessary. Because Sunday was the day of rest for the majority of the population, state courts were convinced that achieving "uniform observance" could only come about by setting aside Sundays.

The Supreme Court's first free exercise decision on the merits also embedded necessity as a limiting principle for the federal Free Exercise Clause. In *Reynolds*, the Court upheld the criminal conviction of George Reynolds for practicing polygamy—a practice he followed based on his beliefs as a member of the Church of Jesus Christ of Latter-day Saints.[40] On the surface, the *Reynolds* case could seem like the Court simply deferring to the legislature's judgment in enacting the law.[41] Indeed, this approach is how many would characterize free exercise cases before language like strict scrutiny and narrow tailoring was used.[42] But the lack of constitutional scrutiny as we know it does not mean early courts did not still require government justification for infringing on constitutional rights.

Like the state cases that preceded it, the Court's decision in *Reynolds* took for granted that guarding against the "evil" effects of polygamy was a matter of government necessity.[43] The *Reynolds* majority explained that the Free Exercise Clause still left Congress "free to reach actions which were in violation of social duties or subversive of good order."[44] Noting that "[p]olygamy has always been odious among the northern and western nations of Europe," the Court had no trouble concluding that Congress's polygamy prohibition was necessary for the public good and thus compliant with the Constitution.[45] Moreover, the historical context provides strong evidence that polygamy was generally considered to be socially destructive. For example, the Republican Party, in its 1856 platform, named polygamy and slavery as the "twin relics of barbarism" to be eradicated from society.[46] Thus, the necessity of

---

40 *Reynolds*, 98 U.S. at 168.

41 *Id.* (explaining how "Congress . . . saw fit to make bigamy a crime").

42 *See* Clark B. Lombardi, Reynolds *Revisited: The Original Meaning of* Reynolds v. United States *and Free Exercise After* Fulton, 75 ALA. L. REV. 1009, 1058–65 (2024) (tracing the view of *Reynolds* as a rational basis review case rather than one involving some measure of judicial scrutiny).

43 *See Reynolds*, 98 U.S. at 168.

44 *Id.* at 164.

45 *Id.* at 164–65.

46 *Republican Party Platform of 1856*, AM. PRESIDENCY PROJECT, https://perma.cc/WH8Z-H528.

limiting polygamy in America formed an undertone of the *Reynolds* opinion that urged the Court to uphold the legislative act.

B. Necessity Moves to the Foreground: *Sherbert* and *Yoder*

Of course, necessity did not stay forever in the background of free exercise jurisprudence. As the Founding Era inheritance of legislative supremacy began to fade,[47] perhaps courts saw their role as interpreters of the Constitution in a new light. This was certainly true in other areas of constitutional law, as signaled by the significant crescendos of judicial power first in *Lochner v. New York*[48] and later in *Brown v. Board of Education*.[49] Whatever the exact reasons, this shift also impacted how necessity was expressed in free exercise jurisprudence. Rather than implicitly relying on necessity, courts began to more explicitly apply a narrow tailoring analysis, requiring a government showing of necessity.

The movement to foreground necessity in the form of narrow tailoring under strict scrutiny was carried forward by two cases. The first was the Supreme Court's 1963 decision in *Sherbert v. Verner*. The case asked whether the Free Exercise Clause required payment of unemployment benefits to a Seventh-Day Adventist who was discharged for refusing to work on Sunday.[50] After acknowledging that South Carolina's actions impeded appellant Adell Sherbert's exercise of religion, the Court asked whether the state had adequately justified this infringement on liberty.[51] To satisfy this inquiry, the Court said that South Carolina must prove that accommodating Sherbert's religious exercise would "endanger[ ] paramount interests."[52] While South Carolina had suggested that allowing accommodations to Saturday worshippers might "dilute the unemployment compensation fund" and hinder employer scheduling, it provided "no proof whatever" to justify its claims.[53] In other words, South Carolina had failed to show that denying exemptions for Saturday worship was truly

---

47 *See, e.g.*, Philip B. Kurland, *The Constitution: The Framers' Intent, the Present and the Future*, 32 ST. LOUIS U. L.J. 17, 20–21 (1987) ("After a long period of legislative hegemony, of what Woodrow Wilson when still a scholar described as *Congressional Government*, power has more and more left the hands of Congress and has been concentrated more and more in the executive and its handmaiden, the judiciary." (emphasis in original)).

48 198 U.S. 45 (1905).

49 347 U.S. 483 (1954).

50 *Sherbert*, 374 U.S. at 399–403.

51 *Id.* at 409–10.

52 *Id.* at 406 (quoting Thomas v. Collins, 323 U.S. 516, 530 (1945)).

53 *Id.* at 407.

necessary to effectuate its unemployment benefits program. Accordingly, the Court determined that the state violated Sherbert's free exercise rights.[54]

Similarly, the Court foregrounded necessity in *Wisconsin v. Yoder*. In that case, the Court found that Wisconsin's compulsory education law violated the free exercise rights of the objecting Amish.[55] Like in *Sherbert*, the *Yoder* Court held that Wisconsin failed to demonstrate that its "interest in compulsory education would be adversely affected by granting an exemption to the Amish."[56] While the state's interest in education was certainly suitable—indeed, compelling[57]—its restriction on religious liberty to achieve this interest was not necessary. The Amish ended formal education following the eighth grade but continued vocational training in their communities. This continued education undermined the state's argument that without compulsory public education, these children would "become burdens on society because of educational shortcomings."[58] Again, the state's infringement on free exercise was invalidated because it was not demonstrably necessary. As in *Sherbert*, the Court in *Yoder* openly embraced the necessity principle through its use of a strict scrutiny analysis.

C. Heuristics for Necessity: *Smith*

Both *Sherbert* and *Yoder* marked the turn to an explicit necessity requirement in free exercise: when government actions burden religious exercise, the government must demonstrate its actions are narrowly tailored to serve a compelling interest.[59] This standard, strict scrutiny, animated free exercise case law until the Court decided *Employment Division v. Smith* in 1990. At issue in *Smith* was an Oregon law that banned controlled substances including peyote—a hallucinogenic drug used in Native American

---

54 *Id.* at 409–10.

55 *Yoder*, 406 U.S. at 234–36.

56 *Id.* at 236 (citing *Sherbert*, 374 U.S. 398).

57 *Id.* at 221.

58 *Id.* at 224.

59 The Court has provided several variations of this test that all point to necessity. *See, e.g.*, *Lee*, 455 U.S. at 257–58 (requiring the government to "justify a limitation on religious liberty by showing that it is essential to accomplish an overriding governmental interest"); Thomas v. Rev. Bd., 450 U.S. 707, 718 (1981) (requiring the state to "justify an inroad on religious liberty by showing that it is the least restrictive means of achieving some compelling state interest").

religious ceremonies.[60] The Court found that enforcing a neutral and "generally applicable prohibition[ ] of socially harmful conduct" like the Oregon law categorically superseded a right to religious freedom.[61] In *Smith*, the Court seemingly sidelined the necessity requirement for a new evaluation of a government burden on religious exercise. This new standard shifted focus to whether the law at issue was a "neutral law of general applicability" rather than a direct evaluation of necessity.[62]

One way to understand neutrality and general applicability is that they serve as heuristics for necessity.[63] As shown in the Court's opinion in *Smith*, a government decision to cover a broad range of conduct (such as when passing a criminal statute) may suggest the importance of enforcing the law uniformly. In other words, the choice to pass neutral and generally applicable laws embodies a sense of necessity that would not exist if the law were riddled with exceptions. This logic suggests that laws containing myriad exceptions are more likely to fail a free exercise necessity test.

Subsequent Supreme Court cases illustrate this heuristic at work. In *Church of the Lukumi Babalu Aye, Inc. v. City of Hialeah*,[64] for example, the Supreme Court used its new *Smith* standard to strike down a city health ordinance prohibiting animal slaughter for sacrifice. Beginning with its neutrality and general applicability inquiry, the Court noted how the city ordinance permitted "almost all killings of animals except for religious sacrifice."[65] Yet, the Court reasoned, many other types of animal killings are allowed under the statute, even though they plausibly have similar effects on the government's interests in public health.[66] As a result, the ordinance was not a neutral and generally applicable law. The city of Hialeah therefore could not justify its disfavor toward religious sacrifice as necessary to public health.[67]

---

60 *Smith*, 494 U.S. at 874.

61 *See id.* at 885 (quoting Lyng v. Nw. Indian Cemetery Protective Ass'n, 485 U.S. 439, 451 (1988)).

62 *Id.* at 879 (quoting *Lee*, 455 U.S. at 263 n.3).

63 *See* Brady Earley, *Responsible Religious Freedom: Factual Scrutiny in Free Exercise Doctrine*, J.L. & RELIGION 2, 7 (Sept. 30, 2024), https://perma.cc/BZX6-EWRA (analyzing cases that suggest "neutrality and general applicability are tools for evaluating government justifications").

64 508 U.S. 520 (1993).

65 *Id.* at 536.

66 *Id.* at 536–38.

67 *Id.* at 546 ("It follows from what we have already said that these ordinances cannot withstand [strict] scrutiny.").

Another example of how neutrality and general applicability are heuristics for necessity is *Fulton v. City of Philadelphia*.[68] In *Fulton*, the Court held that Philadelphia's refusal to renew a foster agency contract with Catholic Social Services (CSS) was unconstitutional. Philadelphia argued that its denial was justified because CSS had refused to place foster children with same-sex couples contrary to the city's nondiscrimination policy. The Court disagreed. While Philadelphia argued that it was necessary to pursue nondiscrimination in foster care, its foster agency contracts explicitly allowed the Commissioner of the Department of Human Services to permit exceptions to this policy "in his/her sole discretion."[69] Such a large exception to the rule, the Court reasoned, meant that the law could not be considered generally applicable. Like in *Lukumi*, the Court once again imported its neutral and general applicability inquiry into its determination of necessity under strict scrutiny: given this broad "system of exceptions" in the contract, the Court concluded that the city had undermined its claim that nondiscrimination "can brook no departures."[70] In effect, it is challenging for the government to argue a law is necessary to achieve a government aim when it creates exemptions for nonreligious interests.

Many commentators have traditionally framed *Smith*'s neutrality and general applicability test as diametrically opposed to strict scrutiny.[71] This approach positions the debate over free exercise between two camps: the *Smith* approach and the strict scrutiny approach. Critics of *Smith* have pointed to the harms of *Smith* on religious minorities or its departure from free exercise

---

68 141 S. Ct. 1868 (2021).

69 *Id.* at 1878 (quotation marks omitted) (quoting Supp. App. to Brief for City Respondents at 16–17, *Fulton*, 141 S. Ct. 1868 (No 19-123)).

70 *Id.* at 1882.

71 *See, e.g.*, Frederick Mark Gedicks, *The Myth of Second-Class Free Exercise*, 70 VILL. L. REV. 1, 9–12, 22–44 (2025) (arguing *Smith* provides parity between free exercise and other First Amendment freedoms); Marc O. De Girolami, *The Death and New Life of Law and Religion*, 13 OXFORD J.L. & RELIGION 16, 33 (2024); Zalman Rothschild, *The Impossibility of Religious Equality*, 125 COLUM. L. REV. 453, 492–524 (2025); Michael Stokes Paulsen, *Freedom for Religion*, 133 YALE L.J.F. 403, 405–17, 425–28 (2023); Christopher C. Lund, *Second-Best Free Exercise*, 91 FORDHAM L. REV. 843, 863–75 (2022); Christopher R. Green, *Citizenship and Solicitude: How to Overrule* Employment Division v. Smith *and* Washington v. Davis, 47 HARV. J.L. & PUB. POL'Y 465, 474–81 (2024); Sherif Girgis, *Defining "Substantial Burdens" on Religion and Other Liberties*, 108 VA. L. REV. 1759, 1793–1804 (2022); Stephanie H. Barclay, *Replacing* Smith, 133 YALE L.J.F. 436, 448–71 (2023) [hereinafter Barclay, *Replacing* Smith]; Gabrielle M. Girgis, *Taming Strict Scrutiny*, 76 FLA. L. REV. 849, 879–904 (2025).

history and tradition.[72] Critics of strict scrutiny in free exercise point to its "unsustainability" due to permitting too many religious exemptions.[73] Across each proposed path forward for free exercise, scholars have highlighted the tension between *Smith* and strict scrutiny.[74]

As argued in this Part, however, *Smith* and strict scrutiny are *not* in direct opposition; they share a common search for the necessity of government action burdening religion. This framing accords with Professor Stephanie Barclay's theory of constitutional rights as protected reasons,[75] which posits that constitutional rights derive protection from placing an evidentiary burden on government officials.[76] The necessity test argued here does the same in cases in which governments burden religious exercise: governments must provide adequate reasons for these burdens' necessity, whether as a compelling interest under strict scrutiny or as an explanation for why the policy is generally applicable and neutral. Of course, this test raises the question of how government officials can satisfy this evidentiary burden.

II. OPERATIONALIZING NECESSITY: DIFFERENCE-IN-DIFFERENCES

Even if necessity lies at the core of free exercise case law, as argued in Part I, how can courts operationalize necessity going forward? A test like strict scrutiny, which places necessity at the foreground, raises the concerns posed by Justice Brett Kavanaugh and others that judges will use these tests to conduct a freewheeling

---

[72] *See* Paulsen, *supra* note 71, at 405–17, 425–28 (offering a theory that relies on protecting freedom for religion based on historical understandings); Lund, *supra* note 71, at 870–75 (discussing the barriers faced by religious minorities and criticizing *Smith* and its progeny).

[73] De Girolami, *supra* note 71, at 30–34 (doubting the practicality of a true strict scrutiny regime in free exercise); *see also* Andrew Koppelman, *The Increasingly Dangerous Variants of the "Most-Favored-Nation" Theory of Religious Liberty*, 108 IOWA L. REV. 2237, 2253 (2023) (discussing how the Supreme Court "distorts the application of strict scrutiny after triggering it"). *See generally* Rothschild, *supra* note 71 (arguing that the Court's current approach to religious equality in free exercise cases is deficient).

[74] *See, e.g.*, Lund, *supra* note 71, at 863 ("[I]f we are going to oversimplify and say *Smith* was just about one thing, . . . *Smith* wanted courts out of the business of balancing governmental interests against religious ones.").

[75] *See generally* Stephanie H. Barclay, *Constitutional Rights as Protected Reasons*, 92 U. CHI. L. REV. 1179 (2025) [hereinafter Barclay, *Constitutional Rights*].

[76] *Id.* at 1240–44.

policy assessment leading to unfettered judicial discretion.[77] Indeed, this was one of the arguments that Justice Antonin Scalia and the majority in *Smith* used to turn away from a direct application of necessity through strict scrutiny.[78] Strict scrutiny, Justice Scalia explained, would enable judges to "weigh the social importance of all laws."[79] This concern is illustrated by cases involving religious exemptions to vaccines, in which one judge may consider vaccines necessary to public health while another may take a different view.[80]

These concerns are ultimately addressable. When it comes to acceptable reasons for government action, courts should use history as a guide.[81] Historically, religious freedom was limited when it posed threats to public safety, peace, or order.[82] Yet questions remain about how to determine whether government action is necessary to achieve one of these historically supported reasons. In free exercise cases, government officials have struggled to produce evidence of necessity, leaving judges to intuit their own answers. Part II.A explores this challenge. Part II.B offers a novel solution applying the difference-in-differences methodology. Using DiD, courts and state lawmakers can carefully review evidence of necessity by cross-evaluating the state's empirical outcomes with other states that have comparable practices. This approach constrains judicial discretion while also enabling courts to conduct a robust necessity inquiry in free exercise cases.

A. Problems with Existing Evidentiary Gaps

Notwithstanding the underlying preservation of necessity in *Smith*'s neutrality and general applicability standard, the doctrine has left many unsatisfied.[83] This may be due to the reality

---

77 *See* Ramirez v. Collier, 142 S. Ct. 1264, 1288 (2022) (Kavanaugh, J., concurring) ("[T]he compelling interest and least restrictive means standards require this Court to make difficult judgments about the strength of the State's interests and whether those interests can be satisfied in other ways that are less restrictive."); J. Joel Alicea & John D. Ohlendorf, *Against the Tiers of Constitutional Scrutiny*, 41 NAT'L AFFS. 72, 73–74, 81 (2019) (denoting such approaches as "quintessentially political").

78 *See Smith*, 494 U.S. at 890.

79 *Id.*

80 *See infra* notes 83–85 and accompanying text.

81 *See* Barclay, *Constitutional Rights*, *supra* note 75, at 1209–14, 1242.

82 *See, e.g.*, *supra* notes 27–29 and accompanying text.

83 *See Fulton*, 141 S. Ct. at 1931 (Gorsuch, J., concurring) ("*Smith* has been criticized since the day it was decided. No fewer than ten Justices—including six sitting Justices—have questioned its fidelity to the Constitution.").

that neutrality and general applicability are often unhelpful heuristics for necessity. For example, consider a COVID-19 vaccine mandate that lacks a religious exemption and is considered generally applicable.[84] While general applicability may indicate a facially strong government interest in public health, it tells us little about whether denying accommodations to religious objectors is *necessary* for the government to achieve its public health aims. Answering this question requires understanding the effect on public health of a vaccine mandate with no religious exemptions. It simultaneously requires an understanding of the effectiveness of alternatives that are less burdensome on religion. As intimated by Justice Neil Gorsuch, "if a State could prove that granting or denying religious exemptions *would* make the difference between achieving a crucial vaccination threshold, it may be that denying exemptions beyond that threshold number could qualify as a narrowly tailored rule necessary to achieve a compelling state interest."[85] Thus, understanding the impact of a particular state law or policy is the core inquiry of necessity under the Free Exercise Clause and integral to the role of judges in constitutional review applying such facts.

Rather than relying on assumptions about a challenged law's potential efficacy, effective approaches to constitutional construction should enable judges to apply the law to the facts before the court.[86] Unfortunately, state actors too often do not offer such facts or arguments, leaving courts to intuit the necessity of the government's actions themselves.[87] This problem is not exclusive to the free exercise context. A survey of Supreme Court cases applying strict scrutiny reveals that a bare majority (53%) of claims even receive briefing from the state actor to justify its actions to

---

[84] Notwithstanding medical exemptions, COVID-19 vaccine mandates have been characterized by courts as laws of general applicability and reviewed under *Smith*'s more deferential standard. *See* Does 1–6 v. Mills, 16 F.4th 20, 32 (1st Cir. 2021); We the Patriots USA, Inc. v. Hochul, 17 F.4th 266, 290 (2d Cir. 2021).

[85] Dr. A v. Hochul, 142 S. Ct. 552, 557 (2021) (mem.) (Gorsuch, J., dissenting from the denial of application for injunctive relief) (emphasis in original).

[86] *See* U.S. CONST. art. III, § 2, cl. 1; *see also* John Inazu, *First Amendment Scrutiny: Realigning First Amendment Doctrine Around Government Interests*, 89 BROOK. L. REV. 1, 27 (2023) (discussing how category distinctions in free exercise jurisprudence "too often mask what is at stake in First Amendment cases").

[87] *See Dr. A*, 142 S. Ct. at 557 (mem.) (Gorsuch, J., dissenting from the denial of application for injunctive relief) (noting that New York failed to advance arguments about whether its laws were narrowly tailored to some critical vaccine threshold).

the Court.[88] In cases applying strict scrutiny to a free exercise issue, the government provided an evidentiary justification in only ten out of twenty-two decided cases (45%).[89]

Some may posit that perhaps amici are presenting this evidence in lieu of party briefing.[90] The data, however, suggest this substitution is not the case. Of the twelve free exercise cases in which state actors did not provide evidentiary justifications, amici were cited as supplying this evidence in only two cases.[91] Often, when the Justices found the evidence of necessity lacking, they simply determined that the state had failed to carry its burden of proof. This was the outcome in *Thomas v. Review Board of Indiana Employment Security Division*[92]: the Court found "no evidence in the record" to show that granting a religious accommodation would lead to an unwieldy stream of additional requests for accommodation and thus found the accommodation to be mandated under the Free Exercise Clause.[93]

Cases in which the religious claimant wins by default due to a lack of government evidence may seem like harmless error. Indeed, this outcome provides protection to claimants that they otherwise might not have received. Yet when the Supreme Court repeatedly decides cases without considering any evidentiary justifications for necessity, as the statistics above indicate, lower

---

88 *See* Thomas R. Lee, Bradley Rebeiro, Dane Thorley & Brady Earley, Law as a Matter of Fact: The Inherent Empiricism of Heightened Scrutiny 11–14 (Aug. 2024) (BYU L. Sch. working paper) (on file with author).

89 *Id.* app. A.

90 *See* Allison Orr Larsen & Neal Devins, *The Amicus Machine*, 102 VA. L. REV. 1901, 1919–24 (2016) (discussing how parties can effectively expand their claims beyond brief word limits through "wrangling" amici).

91 *See* Lee et al., *supra* note 88. The first case is *South Bay United Pentecostal Church v. Newsom*, 141 S. Ct. 716 (2021) (mem.), in which Justices Gorsuch, Clarence Thomas, and Samuel Alito found California's COVID-19 restrictions not necessary. For support, they cited evidence supplied by amici that "California is the only State in the country that has gone so far as to ban *all* indoor religious services." *Id.* at 717 (emphasis in original) (citing Brief Amicus Curiae of the Becket Fund for Religious Liberty at 5–6, *S. Bay United Pentecostal Church*, 141 S. Ct. 716 (No. 20A136)). In the second case—*Does 1–3 v. Mills*, 142 S. Ct. 17 (2021) (order denying application for injunctive relief)—Justice Gorsuch, dissenting from denial of certiorari, acknowledged amicus evidence that denying a religious exemption to Maine's vaccine mandate was not compelling given the existence of such an exemption in comparable states. *Id.* at 21–22 (Gorsuch, J., dissenting) (citing Brief Amicus Curiae of the Becket Fund for Religious Liberty in Support of Applicants and Emergency Application for Writ of Injunction at 13, *Mills*, 142 S. Ct. 17 (No. 20A90)). Note that neither case had the advantage of full briefing at their respective stages, further suggesting amici providing this evidence is an anomalous result.

92 450 U.S. 707 (1981).

93 *Id.* at 719.

court judges may be tempted to cheat on the doctrine by surmising necessity based on history, precedent, or intuition. This bears out what Justice Kavanaugh and others have feared strict scrutiny would do: encourage judges to engage in policymaking.[94]

For instance, consider the Court's decision in *Espinoza v. Montana Department of Revenue*.[95] Montana argued that its constitution's no-aid-to-religion clause—which it used to deny publicly funded tuition assistance to parents sending their children to religious schools—"safeguards the public school system by ensuring that government support is not diverted to private schools."[96] The Court rejected this argument as "fatally underinclusive" because only religious private schools are subject to the no-aid requirement.[97] Importantly, the Court's reasoning for this conclusion involved a subtle empirical assumption: stopping aid to religious schools was not necessary to the vitality of public school system support. But this assumption was simply the Court's intuition about the program, not reasoning grounded in fact. It seems that Montana had evidence to rebut this assumption, given that 94% of funds for private schools went to religious schools after the prohibition was enjoined pending litigation.[98] However, Montana made this argument as a defense against hostility to religion rather than evidence of the law's necessity.[99] While it is hard to know if Montana's evidence would have made a difference, that evidence potentially adds important factual context that the Court's decision ultimately lacked.

Like intuition, the Court has also relied on history and precedent to demonstrate necessity. These tools can be helpful in identifying original understandings of the meaning of free exercise and its corresponding limits.[100] However, history and precedent can also misdirect judicial attention toward inapposite historical facts and away from the present facts of the case.

---

94 *See* United States v. Rahimi, 144 S. Ct. 1889, 1912 (2024) (Kavanaugh, J., concurring) (explaining that judges "impos[ing] their own policy views on the American people" is not consistent with "the properly neutral judicial role").

95 140 S. Ct. 2246 (2020).

96 *Id.* at 2261. Note that while Montana made this assertion, its brief made no efforts to justify this provision as necessary to deny aid to religious schools. *See* Brief of Respondents, *Espinoza*, 140 S. Ct. 2246 (No. 18-1195).

97 *Espinoza*, 140 S. Ct. at 2261.

98 Brief of Respondents at 25, *Espinoza*, 140 S. Ct. 2246 (No. 18-1195).

99 *Id.*

100 *See* Barclay, *Replacing* Smith, *supra* note 71, at 455–61 (discussing how history can provide support for the validity (as opposed to the necessity) of an asserted governmental aim).

In *Braunfeld v. Brown*,[101] the Supreme Court upheld Pennsylvania's criminalization of Sunday retail sales over Abraham Braunfeld's objections as an Orthodox Jew. The Court raised concerns that permitting a religious exemption would "undermine the State's goal of providing a day that . . . eliminates the atmosphere of commercial noise and activity"[102] and would grant some groups "economic advantage over their competitors who must remain closed on that day."[103] This advantage would, in turn, produce enforcement challenges for the state and ultimately frustrate the original design of the rule.[104]

To support this conclusion, the Court pointed to historical debates in the House of Lords and House of Commons prognosticating that Jews would gain an economic advantage from exemptions to Sunday closing laws.[105] But the Court's opinion failed to account for arguments challenging that prediction or whether those predictions were borne out beyond the rhetoric at Westminster.[106] Moreover, it is unclear whether the socioeconomic and religious situation in the United Kingdom in 1936 would accurately map onto 1960s Pennsylvania. This same challenge arises when trying to use precedent where the historical context may be inapposite to the present case. Accordingly, without sufficient evidence from the record,[107] courts will be tempted to search for historical support that invites factual speculations, which ultimately distract from the necessity inquiry.

B. A Solution with Methods: Difference-in-Differences

In response to these systematic evidentiary gaps, this Section offers a plausible way to operationalize necessity in free exercise. This approach is known as difference-in-differences. In recent

---

[101] 366 U.S. 599 (1961).

[102] *Id.* at 608.

[103] *Id.* at 608–09.

[104] *Id.* at 609.

[105] *Id.* at 609 n.6 (citing HC Deb (24 Apr. 1936) (311) col. 492; HL Deb (2 Jul. 1936) (101) col. 430).

[106] For those disputing the economic advantage point, see HL Deb (2 Jul. 1936) (101) cols. 429–30.

[107] *See* Appellants' Brief at *15, *Braunfeld*, 366 U.S. 599 (No. 67) ("Counsel has no way of knowing what percentage of Sabbatarians situated like appellants would continue to suffer the economic loss and what percentage would instead quietly give up the faith that until now has been theirs. But the pressure, the compulsion, is there.").

years, DiD has risen to prominence as a scholarly method to address constitutional questions.[108] The rise of DiD coincides with a broader movement toward more credible methods for estimating causal effects.[109] This movement—known as the credibility revolution—has picked up momentum in the legal space due to the abundance of testable causal assumptions across legal doctrines.[110] For example, the Supreme Court's decision in *New York State Rifle & Pistol Ass'n v. Bruen*[111] featured both Justice Stephen Breyer (in dissent) and Justice Samuel Alito (in concurrence) citing DiD studies to reach separate conclusions.[112] However, there have been no proposals yet to incorporate DiD in the free exercise context.

To calculate a DiD estimate, an analyst makes two comparisons. First, they compare an outcome of interest before and after a policy change in a state that adopted the policy (a treated state) and that same outcome in a state that did not adopt that policy (a control state). Second, they compare the changes in the outcome over time between the treated and control states (thus, "difference-in-differences"). To provide a concrete example, consider two states, State *A* and State *B*. Suppose that State *A* adopted a policy and State *B* did not, and that following the adoption of the policy, State *A* saw outcomes improve by ten units each year

---

[108] *See, e.g.*, Daniel E. Ho & Frederick Schauer, *Testing the Marketplace of Ideas*, 90 N.Y.U. L. REV. 1160, 1203–14 (2015) (using DiD to estimate how restrictions on speech impact the marketplace of ideas); Adam Chilton, Justin Driver, Jonathan S. Masur & Kyle Rozema, *Assessing Affirmative Action's Diversity Rationale*, 122 COLUM. L. REV. 331, 371–73 (2022) (employing DiD and variations of DiD to evidence that diversity can lead to improved academic outcomes in higher education). For an example of DiD's use outside of strictly constitutional law contexts, see John J. Donohue, Abhay Aneja & Kyle D. Weber, *Right-to-Carry Laws and Violent Crime: A Comprehensive Assessment Using Panel Data and a State-Level Synthetic Control Analysis*, 16 J. EMPIRICAL LEGAL STUD. 198, 213–15 (2019) (employing a variation of DiD to study the effect of right-to-carry laws on violent crime).

[109] *See generally* Joshua D. Angrist & Jörn-Steffen Pischke, *The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics*, 24 J. ECON. PERSPS. 3 (2010); JOSHUA D. ANGRIST & JÖRN-STEFFEN PISCHKE, MASTERING 'METRICS: THE PATH FROM CAUSE TO EFFECT (2014).

[110] *See generally* Jeff Lingwall & Michelle Vos, *Causal Narratives and Constitutional Scrutiny*, 52 U. MEM. L. REV. 773 (2022). There is even a guide for applying these methods to legal questions. *See generally* ADAM CHILTON & KYLE ROZEMA, TRIAL BY NUMBERS: A LAWYER'S GUIDE TO STATISTICAL EVIDENCE (2024).

[111] 142 S. Ct. 2111 (2022).

[112] *Compare id.* at 2166 (Breyer, J., dissenting) (citing Donohue et. al, *supra* note 108) (finding "shall-issue" laws increased rates of violent crime in states adopting such laws), *with id.* at 2158 n.1 (Alito, J., concurring) (citing *Effects of Concealed-Carry Laws on Violent Crime*, RAND CORP. (Apr. 22, 2020), https://perma.cc/EK35-ZUBA) (spelling out where shall-issue carry laws are strongly related to increases in violent crime and where the evidence remains inconclusive).

while State *B* only saw outcomes improve by two units each year. To calculate the DiD estimate of the policy's effect, we would take the difference between the change in State *A* (ten units) and the change in State *B* (two units) to yield an estimated effect of an annual eight-unit increase resulting from the policy.

Adopting DiD resolves two potential critiques of a necessity test in free exercise. First, DiD enables necessity to be a workable test across free exercise cases while also encouraging the government to offer evidence-based justifications. DiD is workable because it is an accessible method in a federalist system in which states can typically look to each other for evidence on policy.[113] Indeed, it is quite common for judges to look to use this type of interstate comparison to guide their strict scrutiny inquiry, albeit frequently with less evidentiary basis.[114]

Second, DiD effectuates the free exercise necessity test by asking judges to make determinations based on evidentiary questions rather than open-ended policy preferences. DiD narrows the inquiry from high-level legislative questions like "what is the appropriate policy to ensure traffic safety while still protecting Amish religious practices" to more targeted questions like "does Ohio's crash evidence demonstrate flashing lights improved traffic safety better than less restrictive alternatives." Answering the latter question still involves judicial discretion as judges weigh the evidence before them. But the discretion exercised is about evidence rather than policy.

---

[113] *See* New State Ice Co. v. Liebmann, 285 U.S. 262, 311 (1932) (Brandeis, J., dissenting) ("It is one of the happy incidents of the federal system that a single courageous state may, if its citizens choose, serve as a laboratory."). For additional commentary on how the concept of "laboratories of democracy" functions today, see Charles W. Tyler & Heather K. Gerken, *The Myth of the Laboratories of Democracy*, 122 COLUM. L. REV. 2187, 2204–22 (2022) (discussing the role of third-party networks in policy experimentation) and Gerald S. Dickinson, *The New Laboratories of Democracy*, 1 FORDHAM L. VOTING RTS. & DEMOCRACY F. 261, 266–69 (2023) (arguing the importance of state judicial courts in fostering policy experimentation).

[114] *See, e.g.*, *Yoder*, 406 U.S. at 226 n.15 ("Even today, an eighth grade education fully satisfies the educational requirements of at least six States."); Janus v. Am. Fed'n of State, Cnty., & Mun. Emps., Council 31, 138 S. Ct. 2448, 2466 (2018) (referencing the "federal employment experience" and that of "28 States" to demonstrate that "'labor peace' can readily be achieved 'through means significantly less restrictive of associational freedoms' than the assessment of agency fees." (quoting Harris v. Quinn, 573 U. S. 616, 648–49 (2014))); San Antonio Indep. Sch. Dist. v. Rodriguez, 411 U.S. 1, 27 n.64 (1973) (challenging the assumption of comparative wealth discrimination in Texas schools by reference to studies in California and Kansas); Reeves, Inc. v. Stake, 447 U.S. 429, 444–45 (1980) (undercutting arguments against South Dakota's resident-preference program for concrete distribution as violative of the Commerce Clause by comparing it with a similar program in Maryland that was upheld with even more protectionist features).

Beyond responding to the criticisms of strict scrutiny, DiD is also more feasible for judges to consider than many other empirical techniques. The gold standard for empirical evidence is randomized controlled trials (RCTs) and field experiments, in which subjects are randomly assigned into a treatment or control group and then administered an intervention.[115] Yet these experimental methods come with notable cost and ethical constraints that governments seeking such evidence would have trouble overcoming.[116] Even when a government can muster the political will to fund a new program, rolling it out to only half of the population based on random assignment may raise due process concerns.

Moreover, RCTs and field experiments are often impractical for governments to use during litigation.[117] RCTs and field experiments usually require at least several months to carry out, let alone be approved and funded. The DiD method, by contrast, is particularly well suited to litigation because it relies on existing observational data that the government itself collects as a matter of course. State governments and federal agencies regularly collect and publish data on public health, education, and many other areas as a matter of good governance and political accountability. This practice sets DiD apart from surveys or other methods that

---

[115] *See, e.g.*, Jason Stoughton, *The Bubble-Bursting, Causality-Revealing Awesomeness of Randomized Control Trials*, U.S. NAT'L SCI. FOUND. (Sept. 22, 2022), https://perma.cc/S474-HVAZ; Netta Barak-Corren & Tamir Berkman, *Constitutional Consequences*, 99 N.Y.U. L. REV. 785, 803–22 (2024) (implementing a field experiment and survey analysis to probe a question in free exercise case law); Netta Barak-Corren, Yoav Kan-Tor & Nelson Tebbe, *Examining the Effects of Antidiscrimination Laws on Children in the Foster Care and Adoption Systems*, 19 J. EMPIRICAL LEGAL STUD. 1003, 1013–31 (2022) (using machine learning to examine an empirically driven assumption in free exercise doctrine); Netta Barak-Corren, *Religious Exemptions Increase Discrimination Towards Same-Sex Couples: Evidence from* Masterpiece Cakeshop, 50 J. LEGAL STUD. 75, 83–93 (2021) [hereinafter Barak-Corren, *Religious Exemptions*] (utilizing a field experiment to examine the effect of a religious freedom case); *cf.* Creighton Meland & Stephen Cranney, Measuring and Evaluating Public Responses to Religious Rights Rulings (Jul. 3, 2022) (unpublished manuscript) (available at https://perma.cc/Z2M6-6FCS) (challenging Barak-Corren's experimental evidence).

[116] For a more comprehensive discussion of RCT critiques, see Timothy Ogden, *RCTs in Development Economics, Their Critics and Their Evolution*, *in* RANDOMIZED CONTROL TRIALS IN THE FIELD OF DEVELOPMENT: A CRITICAL PERSPECTIVE 126, 128–31 (Florent Bédécarrats et al. eds., 2020); Angus Deaton & Nancy Cartwright, *Understanding and Misunderstanding Randomized Control Trials*, 210 SOC. SCI. & MED. 2, 17–18 (2018).

[117] *See* Barclay, *Constitutional Rights*, *supra* note 75, at 1279 (noting the "epistemological" bounds imposed on courts and parties by litigation). *But cf.* Chelsey Nelson Photography, LLC v. Louisville/Jefferson Cnty. Metro Gov't, 624 F. Supp. 3d 761, 798 (W.D. Ky. 2022) (declining to admit Professor Netta Barak-Corren's experimental evidence as an expert opinion).

rely on more qualitative observations with more potential for subjective manipulation. Although DiD brings its own problems, such as noisy estimates and *p*-hacking—a common problem in empirical methods generally[118]—it provides tremendous benefit through tangibly justifying the assumptions used to generate policy, in turn leading to greater transparency. This is virtuous in the context of strict scrutiny because a lack of statistical or substantive significance is still relevant to the case—it is an indication that the state has not carried its burden to demonstrate that its actions are necessary.

Of course, DiD is a method of quantitative analysis and therefore raises some common lines of objection. First, there are many things governments care about that are not easily measured.[119] Thus, in many cases, we might expect that DiD will fail to adequately demonstrate the state's interests. Consider, for example, Philadelphia's interest in nondiscrimination in *Fulton*.[120] Nondiscrimination laws, Philadelphia argued, promote expressive values such as fairness and equality. How does a state demonstrate, or a court evaluate, when pursuing such interests justifies infringement on religious liberty? The answer is deceptively simple. All laws, even those that aim to pursue difficult-to-measure expressive values, do so through some form of regulated conduct. In *Fulton*, the policy provided same-sex couples uniform access to foster care. In the example of public accommodations, the measurable variable could be access to a hotel or a restaurant.[121] Measuring access is amenable to a DiD inquiry.[122] And given our modern survey and data capabilities, empirical techniques can plausibly measure expressive values like social stigma.[123] Thus, access and

---

[118] Andrew Gelman & Eric Loken, *The Statistical Crisis in Science*, AM. SCIENTIST (2014), https://perma.cc/E63C-XARA. The *p*-hacking problem arises from researchers reworking their analysis until they generate a statistically significant approach. Noisy estimates occur when the estimated effect may have such a large confidence interval that the true size of the effect is hard to discern.

[119] *See, e.g.*, *infra* text accompanying notes 192–94.

[120] *Fulton*, 141 S. Ct. at 1882.

[121] *See* Heart of Atlanta Motel, Inc. v. United States, 379 U.S. 241, 243 (1964) (involving access to motels); Katzenbach v. McClung, 379 U.S. 294, 300 (1964) (involving access to restaurants).

[122] For example, using cell phone data is one way that researchers can proxy access based on where people live and how far they usually travel. *See* Brady Earley, Data-Driven Accommodations: Testing Religious Exemptions in Markets of Discrimination 21–38 (2025) (working paper) (on file with the author).

[123] *See, e.g.*, Barak-Corren, *Religious Exemptions*, *supra* note 115 (using surveys in a field experiment to understand how discrimination toward same-sex couples changed in

stigma can both serve as proxies for the state's interest in nondiscrimination and offer a concrete approach for justifying the necessity of governmental encroachment on religious exercise.

There is a related reason why the immeasurable interests concern should not preclude using DiD. Recall that the burden is on the government to demonstrate that rejecting a religious accommodation is necessary to serve its interests.[124] Under this burden, courts defer to the government in determining what interests are relevant,[125] but they require a "precise analysis" to prove a challenged state action is needed to achieve those interests.[126] It follows that states choosing to prove abstract or hard-to-quantify interests are probably stacking the deck against themselves. For example, in *Moody v. NetChoice, LLC*,[127] Texas and Florida passed laws limiting social media platforms' ability to engage in content moderation with the aim to "correct the mix of speech that the major social-media platforms present."[128] The Supreme Court remanded the cases back to the lower courts instead of ruling on the constitutionality of the social media laws but left serious doubts about whether the laws were constitutional.[129] After all, how could a state show, or by what standard could a court tell, when the "correct" mix of speech is accomplished? By stating such subjective aims, Texas and Florida could move the goalposts as they pleased, which in turn would allow them to potentially push past constitutional boundaries. This result—courts being pushed by states to expand the ambit of the Free Speech Clause—would be hard to square with the judicial duty to "say what the law is."[130] By the same logic, the existing free exercise jurisprudence encourages government defendants to choose interests that they can demonstrate with precision. It is an advantage of the DiD method

---

response to the Supreme Court's decision in *Masterpiece Cakeshop v. Colorado Civil Rights Commission*, 138 S. Ct. 1719 (2018)).

[124] Kennedy v. Bremerton Sch. Dist., 142 S. Ct. 2407, 2421 (2022) ("If the plaintiff carries [the] burden[ ] [to demonstrate a rights infringement], the focus then shifts to the defendant to show that its actions were nonetheless justified and tailored consistent with the demands of our case law.").

[125] *See* Carson v. Makin, 142 S. Ct. 1987, 1998 (2022) (rejecting the dissenting Justices' assertions of the government's interest in infringing religious exercise and instead sticking to the interests and rationale provided by the state).

[126] *Fulton*, 141 S. Ct. at 1881 (citing Gonzales v. O Centro Espírita Beneficente União do Vegetal, 546 U.S. 418, 430–32 (2006)).

[127] 144 S. Ct. 2383 (2024).

[128] *Id.* at 2407.

[129] *Id.* at 2409.

[130] Marbury v. Madison, 5 U.S. (1 Cranch) 137, 177 (1803).

that it provides further encouragement for governments to pursue quantifiable, rather than abstract, policy goals.

A final objection takes the opposite tack: even though we can measure the efficacy of many policies with good data, courts are not necessarily well positioned to evaluate such data. This objection is unfounded. Courts have shown a capacity for evaluating even more technical evidence than DiD,[131] including in the strict scrutiny context.[132] One example is the Supreme Court's handling of algorithmic evidence in racial gerrymandering.[133] The Court's evaluation of this evidence has drawn approval from scholars and experts in the methods used in these cases.[134] Certainly there are counterexamples: the Court has received criticism for its mishandling of empirical evidence.[135] But even when dueling statistical studies are placed before the court, a world where judges are evaluating evidence is better than one where they are forced to imagine it.[136] Judges are better positioned to faithfully fulfill their duty to "say what the law is" when they know what the facts are.

III. EVALUATING NECESSITY: A DID ANALYSIS OF AMISH TRAFFIC SAFETY

The previous Part outlined why DiD should be used to demonstrate necessity in free exercise cases. This Part applies DiD to an emerging free exercise conflict over recent changes in Ohio traffic laws that require Amish buggies to display flashing

---

[131] *See, e.g.*, Brown Shoe Co. v. United States, 370 U.S. 294, 325 (1962) (setting forth the test for market definition to adjudicate antitrust claims, which invites courts to evaluate "cross-elasticity of demand" and other factors).

[132] *See* Students for Fair Admissions, Inc. v. President and Fellows of Harvard Coll., 397 F. Supp. 3d 126, 158–77 (D. Mass. 2019) (ruling between competing statistical experts on the question of discrimination, or lack thereof, in college admissions), *rev'd*, 143 S. Ct. 2141 (2023); *see also* Edward K. Cheng, *The Myth of the Generalist Judge*, 61 STAN. L. REV. 519, 540–44 (2008) (finding that federal appellate judges will often be assigned opinions according to specialization which can also allow for a judge with relevant expertise to review more technical evidence).

[133] *See* Allen v. Milligan, 143 S. Ct. 1487, 1512–14 (2023).

[134] *See* Brief for Amici Curiae Nicholas O. Stephanopoulos and Jowei Chen in Support of Appellees at 19, Alexander v. S.C. State Conf. of the NAACP, 144 S. Ct. 1221 (2024) (No. 22-807) (describing the Court's analysis of such of evidence in *Allen* as "apt" and used to "properly reject[ ]" certain computational-redistricting evidence).

[135] *See* Emmet J. Bondurant, Rucho v. Common Cause—*A Critique*, 70 EMORY L.J. 1049, 1086–88 (2021) (criticizing Chief Justice John Roberts for ignoring outcomes of elections in gerrymandered states and expert testimony in an election gerrymandering case).

[136] *Compare, e.g.*, Brown v. Ent. Merchs. Ass'n, 564 U.S. 786, 799–804 (2011) (evaluating evidence from the state's experts about whether violent video games cause harm to minors), *with id.* at 817–20 (Alito, J., concurring) (providing independently researched evidence on the violence of video games).

lights for safety. This case study is a prime example of the need for evidence and the valuable possibilities of DiD in free exercise cases. It also illustrates the role that lawmakers play in constitutional construction when offering evidence of necessity to meet their burden of proof.[137] Part III.A sets forth the legal context and existing case law. Part III.B then conducts a DiD analysis to evaluate the necessity of the light requirement to achieve Ohio's interest in traffic safety. Finally, Part III.C provides a discussion of the results and responds to potential critiques.

A. Legal Background

In 2022, Ohio's General Assembly passed a law requiring all horse-drawn buggies to display a "yellow flashing lamp" at all times.[138] This law was created after several counties saw an increase in buggy accidents in recent years.[139] The Swartzentruber community in Ohio—an especially conservative branch of the Old Order Amish who object to using many forms of modern technology on religious grounds—protested the law.[140] After the law passed, members of the Swartzentruber in Ashland County refused to comply with the electric light requirement.[141] These individuals were charged with a misdemeanor in the Ashland Municipal Court and fined $50.[142] The Swartzentruber refused to pay the fine and argued that the Free Exercise Clause prohibited the state from enforcing Ohio H.B. 30 against them.[143] They subsequently moved for a preliminary injunction in a neighboring county to "immediately

---

[137] *See* Kennedy v. Bremerton Sch. Dist., 142 S. Ct. 2407, 2426 (2022) (explaining that government bears the proof to show that its actions burdening religion were narrowly tailored to achieve a compelling state interest).

[138] OHIO REV. CODE ANN. § 4513.114 (West 2025).

[139] Kevin Lynch, *"Buggy Bill" Passes; Law Requires Flashing Lights on Buggies*, THE DAILY RECORD (June 19, 2022), https://perma.cc/BB9Y-KEFR; *see also* Jordan Laird, *Amish Buggy Collisions with Cars, Trucks: How Can Ohio Roads Be Safer?*, COLUMBUS DISPATCH (July 15, 2020), https://perma.cc/7TCS-PQHP ("There were more than 700 reported crashes involving a horse-drawn vehicle on Ohio's state routes between 2009 and 2018.").

[140] *See* Kelsey Osgood, *In a Remote Corner of Ohio, a Traffic Law Brings Harvard to the Aid of the Amish*, RELIGION NEWS SERV. (Aug. 21, 2024), https://perma.cc/ADD9-KXEC. The Swartzentruber "use more limited technology, dress more plainly, and typically have a lower standard of living than more progressive Amish." Erik Wesner, *Swartzentruber Amish: The "Hardest-Core" Subgroup (Here's Why)*, AMISH AM. (last updated Jul. 2021), https://perma.cc/67A2-DLEB.

[141] Osgood, *supra* note 140.

[142] *5 Amish Men Plead "No Contest" to Violating Buggy Law in Ashland, Refuse to Pay Fines*, ASHLAND SOURCE (Oct. 19, 2022), https://perma.cc/67YA-VHDV.

[143] Osgood, *supra* note 140.

halt enforcement of the law statewide"[144] and have so far prevailed in lower courts.[145]

The Ohio law is only the most recent example of legislation requiring safety mechanisms on horse-drawn buggies.[146] In fact, Michigan, Minnesota, Wisconsin, Pennsylvania, and Kentucky have all passed similar legislation and faced similar challenges. The earliest challenge came against a Michigan law that required an orange, fluorescent triangle on all slow-moving vehicles (SMVs).[147] The Michigan Court of Appeals invalidated the law as a violation of the Free Exercise Clause because the state failed to show that the requirement was necessary and that it was unable to allow alternative means to comply.[148]

A similar law was challenged by Amish residents of Minnesota soon after the Supreme Court decided *Smith*.[149] The Minnesota Supreme Court relied on the state constitution's free exercise protection and analyzed the case under strict scrutiny.[150] Parallel to the result in Michigan, Minnesota found that the state had failed to meet its burden of showing that the law—without exceptions—was necessary.[151]

For yet another law requiring orange triangles for SMVs,[152] Wisconsin was taken to court by members of the state's Old Order Amish over the law's religious freedom ramifications. After reviewing expert witness testimony supporting a less restrictive alternative that used dull white tape, the Wisconsin Supreme Court, using a strict scrutiny analysis, ruled the traffic law unconstitutional under the state constitution.[153]

In Pennsylvania, a traffic law requires an orange, fluorescent triangle on all SMVs.[154] The law was likewise contested by the Swartzentruber Amish. In 2002, a Pennsylvania county court

---

144 *Id.*

145 *See* Olivia Klein, *Aiding the Amish: Harvard Law Students Secure Early Win in Challenge to Ohio's Buggy Light Law*, HARV. L. TODAY (Feb. 4, 2025), https://perma.cc/26TH-U9A8.

146 Notably, New York is now considering a similar law to Ohio's. John Whittaker, *Amish Buggy Headlight, Taillight Bill Introduced in Legislature*, THE POST-JOURNAL (Jan. 3, 2024), https://perma.cc/SPW3-8KFC.

147 MICH. COMP. LAWS § 257.688 (2025).

148 People v. Swartzentruber, 429 N.W.2d 225, 229 (Mich. Ct. App. 1988) (withholding judgment about the validity of the Amish's proposed alternative of reflective tape).

149 MINN. STAT. § 169.522 (2024).

150 State v. Hershberger, 462 N.W.2d 393, 396–97 (Minn. 1990).

151 *Id.* at 399.

152 WIS. STAT. § 347.245 (2025).

153 State v. Miller, 549 N.W.2d 235, 242 (Wis. 1996).

154 75 PA. CONS. STAT. § 4529 (2025).

ruled that the state's interest in safety could not allow a gray tape alternative that would comport with Amish beliefs.[155] In other words, Pennsylvania was able to meet the burden of strict scrutiny by showing that no less restrictive alternatives were available that would still accomplish the state's goals in traffic safety. As in the Wisconsin case, an expert witness stated that a gray tape alternative could be "better than the retroreflectivity" of the state's SMV emblem in some circumstances.[156] But in the Pennsylvania case, the court determined that the gray tape was not as readily identifiable as the orange SMV emblem that was used in forty-one states at the time.[157] Therefore, the Pennsylvania court sided with the state and rejected the religious freedom challenge.

In a fifth and final example, the Amish challenged a Kentucky law requiring an orange triangle for all SMVs.[158] The Kentucky Supreme Court upheld the law because it was rationally related to the goal of traffic safety—the default standard of review for neutral and generally applicable laws.[159] In response, the Kentucky legislature created a religious accommodation that allowed the Amish to use white or silver reflective tape.[160]

The saga of Amish cases across these five states represents a challenge within free exercise doctrine. In sum, Amish free exercise litigation over buggy regulations has ultimately produced three free exercise wins and two losses. In reaching these results, these five state courts have offered four different reasons to reach these respective outcomes: (1) the state failed to show the law was necessary without exceptions under strict scrutiny (Michigan and Minnesota); (2) a less restrictive alternative was available under strict scrutiny (Wisconsin); (3) no less restrictive alternatives were available under strict scrutiny (Pennsylvania); and (4) the law was neutral and generally applicable, and it satisfied rational basis review (Kentucky). While the laws did arise in different constitutional contexts, they each share a fundamental tension between protecting public safety and the right to free exercise. Taking these cases as a guide, it is far from clear how Ohio courts

---

155 Commonwealth v. Miller, 57 Pa. D. & C.4th 11, 23–24 (Pa. Ct. Com. Pl. Cambria Cnty. 2002).

156 *Id.* at 22.

157 *Id.* at 22–23.

158 KY. REV. STAT. ANN. § 189.820 (West 2025).

159 Gingerich v. Commonwealth, 382 S.W.3d 835, 839–44 (Ky. 2012).

160 Roger Alford, *Amish Buggy Safety Bill Wins Final Passage in Kentucky*, INS. J. (Mar. 29, 2012), https://perma.cc/2NQY-UASK.

would appropriately apply the "narrowly tailored" prong of the Ohio Constitution's strict scrutiny test.[161]

This doctrinal divergence over similar buggy light laws suggests the need for more evidence. Without more evidence informing individual disputes, trying to search for necessity in the government's actions becomes a challenging task. But even when that evidence is brought, how would a court determine whether lights on buggies are necessary to public safety? It is certainly true that more lights are likely to make a buggy more visible (and expert witnesses can so testify), but is that additional visibility actually necessary to enable traffic safety? Existing commentary has so far failed to develop an answer to the Amish dilemma.[162] And without a clear sense of methodological grounding, courts will essentially perform a free-floating analysis that is largely independent of the empirical realities in place. Finally, as already noted, the Supreme Court has unanimously held that narrow tailoring requires a "precise analysis."[163] How are judges supposed to faithfully conduct this precise analysis in a reliable way?

B. DiD Analysis

A DiD approach addresses these concerns and provides a plausible solution for demonstrating and evaluating the necessity of Ohio's buggy law. This begins with precisely identifying the state interests and chosen means to accomplish those interests. Here, Ohio chose to adopt a law requiring flashing lights on buggies to reduce buggy-related crashes[164] instead of continuing to

---

[161] *See* Humphrey v. Lane, 728 N.E.2d 1039, 1043 (Ohio 2000) (interpreting the state constitution to require strict scrutiny for free exercise cases).

[162] For scholars addressing the Amish and traffic safety, see Daniel A. Crane, *Beyond RFRA: Free Exercise of Religion Comes of Age in the State Courts*, 10 ST. THOMAS L. REV. 235, 254–55 (1998); David E. Steinberg, *Gardening at Night: Religion and Choice*, 74 NOTRE DAME L. REV. 987, 992–93 (1999) (book review); So Chun, Comment, *A Decade After* Smith*: An Examination of the New York Court of Appeals' Stance on the Free Exercise of Religion in Relation to Minnesota, Washington, and California*, 63 ALBANY L. REV. 1305, 1318–20 (2000). The existing literature focuses almost exclusively on two Amish buggy decisions largely because of their temporal proximity to the Supreme Court's decision in *Smith*.

[163] *Fulton*, 141 S. Ct. at 1881 (citing Gonzales v. O Centro Espírita Beneficente União do Vegetal, 546 U.S. 418, 430–32 (2006)).

[164] *See* Lynch, *supra* note 139 (quoting a bill sponsor who stated that "[t]he objective is to reduce crashes").

rely on reflective tape or symbols.[165] By contrast, both Michigan[166] and Kentucky[167] continued to use these reflective materials to promote traffic safety during this time.

The question under a free exercise challenge to Ohio's law is whether a flashing light requirement is necessary to reduce buggy-related crashes. Put another way, if Ohio had followed Michigan's or Kentucky's approach, how would that approach affect buggy-related crashes (if at all)? In this way, Michigan and Kentucky serve as a control group for the DiD analysis. If Ohio's more restrictive law is necessary for traffic safety, we would expect to see a reduction in buggy related crashes relative to our less restrictive control group of Michigan and Kentucky.

The data from each of the three states used in the analysis are buggy-related crash totals at the month level. In each instance, the data were gathered from a state or state-sponsored agency that collected and classified crash data for the entire state.[168] The final dataset included twenty-four months of data from October 2021 through September 2023: twelve months of data before Ohio's buggy light law went into effect in October 2022 and twelve months of data after the effective date.[169] Just over a month after the law took effect, the Amish were being arraigned with criminal charges in Ohio municipal court.[170]

---

165 *See* Tyler Buchanan, *House Agrees to Safer Buggy Traffic Standards, Again Rejects Masks*, OHIO CAP. J. (Dec. 9, 2020), https://perma.cc/ZK5N-V52Y ("Under current law, animal-drawn vehicles such as Amish buggies must display one of two things when traveling during the daylight hours: reflective tape or a 'slow-moving vehicle' emblem.").

166 *Supra* notes 147–48.

167 *Supra* note 160.

168 For Ohio data, see *Crash Dashboard*, OHIO STATE HIGHWAY PATROL (updated weekly), https://statepatrol.ohio.gov/dashboards-statistics/ostats-dashboards/crash-dashboard. The Michigan dataset was obtained through assistance from the Michigan Transportation Research Institute and the Office of Highway Safety Planning. *See Michigan Traffic Crash Facts*, UNIV. OF MICH. TRANSP. RSCH. INST., https://perma.cc/Q4Q2-V567. The Kentucky data were obtained through the state police website using the following query: Person Type Code = "Collision – ANIMAL-DRAWN/RIDDEN", Collision Date After August 30, 2021, Collision Date Before September 1, 2023. *Collision Data*, KY. STATE POLICE (updated daily), http://crashinformationky.org/AdvancedSearch.

169 Andrew J. Tobias, *Gov. Mike DeWine Signs Bill Requiring Flashing Lights for Amish Buggies*, CLEVELAND.COM (June 1, 2022), https://perma.cc/P63U-5HPV. Although the law was passed on June 1 and set to take effect on August 31, 2022, the Ohio State Patrol added an extra one-month grace period to "educate the [Amish] community about the dangerous conditions created when buggies share the roads with cars and trucks." Jack Shea, *NE Ohio Officers Enforcing New Law for Mandatory Lights on Amish Buggies*, FOX8 NEWS (Oct. 11, 2022), https://perma.cc/NJ56-KFTH.

170 *See 5 Amish Men Plead "No Contest"*, *supra* note 142.

The relatively brief time window is necessary because the Ohio law itself was only adopted recently and thus limits the availability of post-treatment data. But using this brief window also provides its own advantages. Shorter time windows prevent related legal changes from weakening the integrity of the design. For example, on October 1, 2023, Ohio began to enforce a law imposing new penalties for distracted driving.[171] Michigan passed a similar law that took effect on June 30, 2023.[172] Lengthening the window of analysis into the past or the future thus increases the risk of capturing effects of other laws. By limiting the dataset from October 2021 to September 2023 in these three states, effects from other legal changes with potentially direct effects on buggy-related crashes are limited.[173]

Importantly, the DiD approach involves several assumptions. The most basic assumption involves whether the control group (Kentucky and Michigan) is a good proxy for what would have happened in Ohio absent the legal change. A common way to probe the validity of this assumption is by comparing the trends in the treatment and control groups prior to the law change. If Ohio's crash trends looked like Kentucky and Michigan's prior to the law change in October 2022, it is plausible to think Ohio's trends would have continued like Kentucky and Michigan absent a legal change. As shown below in Figures 1 and 2, the trends in Ohio are roughly (though not perfectly) parallel to the average of the trends in Michigan and Kentucky prior to October 2022. If the parallel pre-trends assumption is valid, we would expect relevant differences in driving between Ohio and the control states (e.g., state driving culture, weather conditions, etc.) would be fully accounted for in the pretreatment crash data. Thus, the roughly parallel trends here provide evidence that Michigan and Kentucky

---

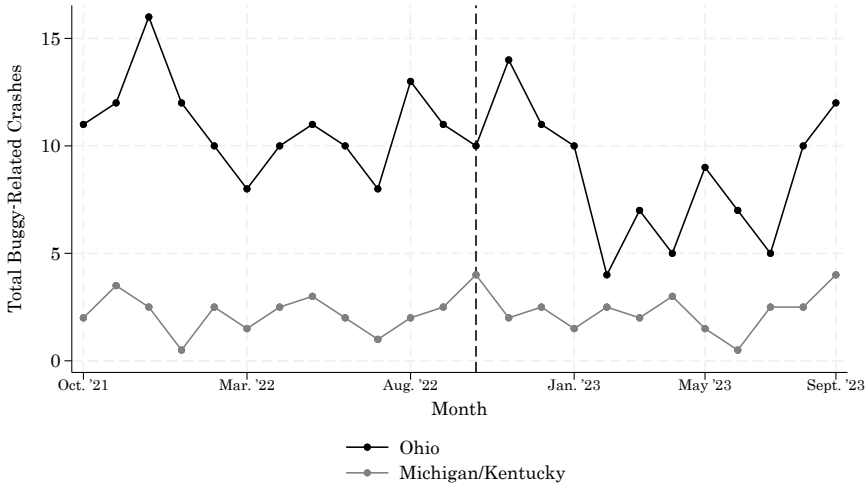
171 Sarah Donaldson, *Ohio's State Agencies Say Data Shows Distracted Driving Law is Working*, THE STATEHOUSE NEWS BUREAU (Oct. 16, 2024), https://perma.cc/EAH3-VVSQ.

172 Cassidy Johncox, *Michigan's New Hands-Free Driving Law Takes Effect Today: What Drivers Should Know*, CLICKONDETROIT (June 30, 2023), https://perma.cc/S3D5-3VRV.

173 Kentucky passed laws relating to testing for blood alcohol content that did not take effect until the end of June 2023. *New Traffic and Impaired Driving Laws—2023 Ky Legislative Session*, KENTUCKY TRAFFIC SAFETY (Apr. 11, 2023), https://perma.cc/TE7A-Y39N. To assuage concerns that this, and Michigan's distracted driving law, may have impacted the last three months of the analysis, running the DiD with an eighteen-month window (January 2021 to June 2023) produces similar results to the twenty-four-month DiD. There was also an unrelated law passed in Michigan in October 2021 that removes license suspensions, but the scope only pertains to minor traffic violations (arguably much different than the crashes dealt with here). *See New Traffic Laws in Michigan: What Drivers Should Know!*, DETROIT JUST. CTR. (Oct. 5, 2021), https://perma.cc/HH42-5NXC.

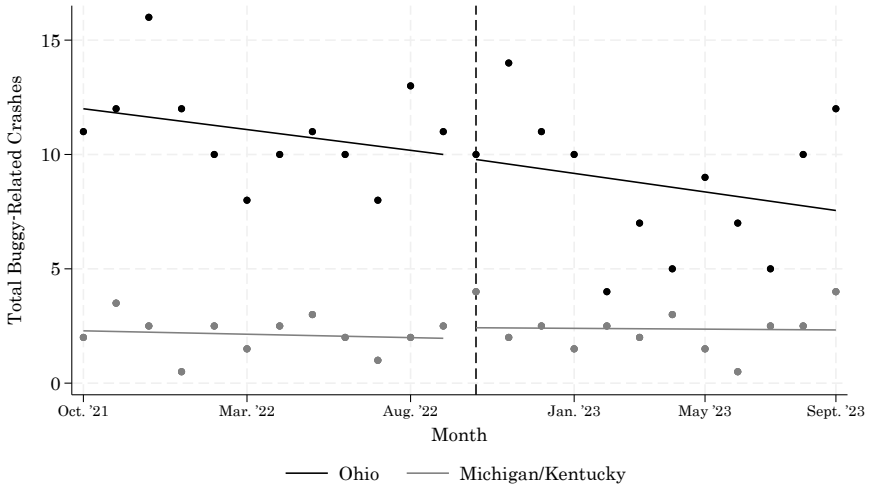
are plausible controls for Ohio. Though more sophisticated methodologies could probe this question with greater precision, this analysis attempts to present a simpler version that would be more approachable to state officials and judges.[174]

FIGURE 1: MONTH-LEVEL CRASH COMPARISON



[174] For those interested in greater sophistication, see Alberto Abadie, Alexis Diamond & Jens Hainmueller, *Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program*, 105 J. AM. STAT. ASS'N 493, 494 (2010) and Dmitry Arkhangelsky, Susan Athey, David A. Hirshberg, Guido W. Imbens & Stefan Wager, *Synthetic Difference-in-Differences*, 111 AM. ECON. REV. 4088, 4089 (2021) (proposing a synthetic DiD to "combine[ ] attractive features of both" DiD and a synthetically created control group using a weighted average of existing control observations).

FIGURE 2: PRE– AND POST–LAW CHANGE LINES OF BEST FIT



The summary statistics of the DiD analysis are presented in Table 1 below. In Ohio, the total number of crashes in the year following the buggy light law's enactment decreased by 28 crashes, an average of 2.75 fewer crashes per month. This change was a 21% decline; in other words, we expect about one fewer crash for every five from before the law was passed. By contrast, the total number of crashes in Michigan and Kentucky stayed roughly the same: the total crashes increased by 8 in the following year between the two states, representing an average increase of 0.67 crashes per month or about a 16% increase.[175] Further, both Kentucky and Michigan exhibited similar trends, so the average does not represent a middle ground between two extremes.

Table B in the Appendix presents the results of the DiD analysis. The point estimate −2.583 represents the average marginal effect of Ohio's flashing light law on buggy crashes, meaning nearly three fewer crashes per month on average. This translates to a 23% decline in buggy crashes on average. Moreover, this result is statistically significant, meaning that it would occur less than 5% of the time if the analysis had generated results using random samples of data.[176]

---

[175] The numbers in Table 1 are rounded to the nearest hundredth, but the full data is available in the Appendix, Table A. The code used to run the DiD analysis is available at the following link: https://github.com/bradyearley/Necessity-in-Free-Exercise/tree/main.

[176] The *t*-statistics for each point estimate in the DiD regression are available in the Appendix, Table B.

TABLE 1: SUMMARY STATISTICS OF BUGGY-RELATED CRASHES

State	Time Period	Crash Total	Crash Monthly Average	Change in Total Crashes
Ohio	Oct. 2021 – Sept. 2022	132	11	−21%
Ohio	Oct. 2022 – Sept. 2023	104	8.67	
Mich./Ky.	Oct. 2021 – Sept. 2022	51	4.25	+16%
Mich./Ky.	Oct. 2022 – Sept. 2023	59	4.92	

C. Discussion of Results

With these results in hand, how could a court use this evidence in its narrow tailoring analysis? Ohio has the burden of proof to demonstrate the necessity of its law. Meeting this burden requires showing that the buggy law is narrowly tailored to a compelling state interest in traffic safety.[177] Ohio has a strong case to do so. The DiD analysis evinces how adopting less restrictive alternatives would undermine Ohio's aim of reducing buggy-related crashes. If Ohio were to continue requiring only reflective tape and symbols on buggies, this analysis anticipates that buggy-related crash levels would likely stay the same. In sum, Ohio has

[177] *See Fulton*, 141 S. Ct. at 1881–82. In reality, the Supreme Court has been somewhat unclear about whether strict scrutiny constitutes a single test or if the government must independently show a compelling interest in addition to narrow tailoring. *See, e.g.*, *Lukumi*, 508 U.S. at 546–47 ("Where government restricts only conduct protected by the First Amendment and fails to enact feasible measures to restrict other conduct producing substantial harm or alleged harm of the same sort, the interest given in justification of the restriction is not compelling."). When the Supreme Court has treated them like separate components, the compelling interest is often assumed, and the analysis proceeds to narrow tailoring. *See, e.g.*, Burwell v. Hobby Lobby Stores, Inc., 573 U.S. 682, 691–92 (2014) ("Under RFRA, a Government action that imposes a substantial burden on religious exercise must serve a compelling government interest, and we assume that the HHS regulations satisfy this requirement.").

empirical support for the necessity of its buggy light law because less restrictive alternatives will "put [Ohio's] goals at risk."[178]

This result is further bolstered by examining how the Supreme Court has previously evaluated statistical evidence under constitutional scrutiny. In past decisions, the Court has placed greater weight on data that can compare a constitutionally challenged state practice to the actions of a counterfactual control group of states.

For example, in *Craig v. Boren*,[179] the Supreme Court rejected Oklahoma's gender-based discrimination on the sale of "3.2% beer."[180] Oklahoma offered several statistics to argue that its restrictions on sales to males under 21 and females under 18 "closely serve[d] to achieve" its goal of traffic safety.[181] In rejecting Oklahoma's policy, the Court pointed out that Oklahoma's evidence was "lacking in controls" sufficient to compare the effectiveness of the more stringent age rule for men against that of the less stringent rule for women.[182] Thus, *Craig* exhibits the Court's acknowledgment and respect for the type of evidence this Comment advocates using.

Of course, there are still many ways that an Amish claimant might rebut the evidence presented.[183] For example, it may be that Kentucky and Michigan are poor control groups for Ohio because their Amish populations are quite different—Ohio's Amish population sits near 80,000 whereas Michigan's and Kentucky's Amish populations are both close to 15,000.[184] This difference in Amish population size, which is disproportionate to the difference in the total populations of these three states, may suggest other unobserved differences between Ohio's Amish population and that of the control states driving the difference in buggy crashes. Perhaps Pennsylvania—another state with around 80,000 Amish

---

178 *See Fulton*, 141 S. Ct. at 1881–82.

179 429 U.S. 190 (1976).

180 *Id.* at 204.

181 *Id.* at 200–04.

182 *Id.* at 202 n.14.

183 Not discussed here, but also important, are qualitative data arguments that should be weighed against these qualitative findings. *See, e.g.*, Corey Anderson, *Horse and Buggy Crash Study II: Overstretching the Slow-Moving Vehicle Emblem's Abilities: Lessons from the Swartzentruber Amish*, 2 J. AMISH & PLAIN ANABAPTIST STUD. 100, 111–12 (2014) (discussing the potential for a "moth effect" where lights may attract drivers toward buggies at night and contribute to crashes).

184 *See Amish Population, 2023*, YOUNG CTR. FOR ANABAPTIST & PIETIST STUD., ELIZABETHTOWN COLL., https://perma.cc/9BRS-FRKL.

people—would be a better comparator.[185] Yet Ohio could point out that in 2022, Pennsylvania removed active and pending driver's license suspensions for drug and other violations.[186] Due to the potential impact this change could have on Pennsylvania's crash data, it is likely a poor control for understanding the effects of the Ohio law enacted the same year.[187] Ohio could also counter the population argument by showing that the same results hold when looking at crashes per capita in the Amish population.[188] Alternatively, the Amish may point out that granting exemptions to the Swartzentruber—a relatively small sect of the Old Order Amish—would produce negligible effects on Ohio's interest in Amish traffic safety generally. Ohio could rejoin this assertion by suggesting that allowing this exemption for the Swartzentruber may encourage other Amish groups to follow suit and eventually render the exempted class so big as to negate the rule.

As these and other objections show, a strong empirical finding will rarely be enough to completely resolve a free exercise case. However, as emphasized throughout this Comment, the goal is less about creating certainty in judicial reasoning and more about adopting a method that enhances accountability. The potential back-and-forth between litigants is exactly the intended benefit of DiD—it produces a common empirical foundation upon which legal arguments can be made by both parties. Judges resolving free exercise disputes are presented with a daunting task: determining whether government action infringing on religious exercise is necessary. The current test under *Smith* attempts to avoid necessity by using neutrality and general applicability as a shortcut. Although this shortcut is misguided, as explained in Part I, judges worry that relying on strict scrutiny will inevitably lead to free-floating balancing tests driven merely by judicial preferences.

The good news is that parties can constrain judicial discretion under free exercise scrutiny by offering better evidence with DiD. While judges will still certainly disagree about the ultimate outcome, they can do so in a precise and factually grounded way. Moreover, under this approach, government officials have a much

---

185 *Id.*

186 *See* 75 PA. CONS. STAT. § 1532(b), (d) (2025).

187 *See Focusing License Suspension on Roadway Safety Offenses*, NAT'L SAFETY COUNCIL 1 (2020), https://www.nsc.org/getattachment/c6c96a66-ff60-484d-9149-0e1cfa87ba6a/t-focusing-license-suspension-roadway-safety-offenses ("20% of traffic fatalities involve an unlicensed driver or one with a suspended license.").

188 Running this analysis achieved similar results in this case.

stronger incentive to legislate and make policy with an eye toward pursuing less restrictive alternatives consistent with their interests. State actors in a DiD regime know that they will face challenges in the courts if they do not have an empirically justifiable necessity guiding their policy decision. Even as this approach constrains judicial discretion, it does not eliminate the judge's role. Faced with evidence from DiD, judges may query whether the state used the correct comparison states or whether the results are substantively significant enough to show necessity. These questions still require judicial discretion, but that discretion will be focused on whether the evidentiary burdens are met in the case at hand—not on whether the policy at issue comports with their preferences.

IV. BEYOND BUGGIES: DID ACROSS FREE EXERCISE

While the Ohio buggy law presents a clear and present case study of DiD, the method's utility reaches beyond that controversy. Three cases before the Supreme Court in 2025 implicated the Free Exercise Clause.[189] These cases provide a useful illustration for filling out the discussion of DiD by showing how future empirical studies may utilize this approach in different ways. Moreover, these cases are also helpful for understanding the limits of DiD and what circumstances may make DiD less compelling.

A. *Mahmoud v. Taylor*

The first case that helps illustrate how DiD could hypothetically be used is *Mahmoud v. Taylor*[190]—a challenge to a Maryland school district's required instruction on gender and sexuality for elementary school children. In denying a preliminary injunction, the Fourth Circuit affirmed the district court's determination that parents' religious objections to "LGBTQ-[i]nclusive [b]ooks" in their children's required curriculum failed to "demonstrate a cognizable burden" on their free exercise.[191] Because the Supreme

---

[189] *See generally* Cath. Charities Bureau, Inc. v. Wis. Lab. & Indus. Rev. Comm'n, 145 S. Ct. 1583 (2025); Okla. Statewide Charter Sch. Bd. v. Drummond *ex rel.* Oklahoma, 145 S. Ct. 1381 (2025) (mem.); Mahmoud v. Taylor, 145 S. Ct. 2332 (2025). Because this Comment was written before the Supreme Court resolved these cases, most of the following discussion refers to lower courts' dispositions of the cases. The information presented in this Comment was reviewed after the cases were decided to ensure its continued accuracy.

[190] 145 S. Ct. 2332 (2025).

[191] Mahmoud v. McKnight, 102 F.4th 191, 197, 202–03 (4th Cir. 2024), *rev'd sub nom.* Mahmoud v. Taylor, 145 S. Ct. 2332 (2025).

Court agreed with petitioners that there was a cognizable burden, the next question turned on whether such a burden was necessary to the state's interests.

At first blush, this case may seem ill-suited to a DiD inquiry given the broad interests in equal treatment that seem to underlie the school board's position and the aforementioned difficulty with quantifying expressive values.[192] But the school district can frame the necessity inquiry in a much more precise way consistent with free exercise precedent.[193] For example, the school district may assert that the required instruction is necessary to preempt high bullying rates of LGBTQ students in the district. They could demonstrate this necessity by comparing the bullying rates over time in school districts that adopted LGBTQ-inclusive curricula and did not allow religious opt-outs with school districts that adopted similar curricula but allowed for religious opt-outs. This DiD approach would then allow the court to evaluate the necessity of the inclusive books policy.[194] Of course, the challenge for the school district is that presenting its interest in such a falsifiable way would only make sense if the data supported their policy. But that is precisely the point of embracing DiD analysis: encouraging states to take a careful look at their policies and bring forth the best evidence to justify that the policies are actually necessary.

Some may consider this possible use of DiD an overly optimistic view of its ability to evaluate necessity in this case. For instance, we may question whether a simple curriculum change, on its own, will be enough to generate a statistically detectable effect on bullying rates. Many small policies, in addition to required instruction with LGBTQ-inclusive books, may cumulatively result in a significant change in outcomes such as bullying. In other words, some policies may be necessary but not sufficient to further a compelling state interest. If that is true, DiD may be too rigid or shortsighted to help determine whether the school district's book policy is necessary to further its aims.

While this argument can be persuasive, there are reasons to resist it. DiD has thus far been applied by looking for a positive change in the challenged state's outcomes. But DiD showing

---

192 *See supra* notes 119–30 and accompanying text.

193 *See Fulton*, 141 S. Ct. at 1881 (citing Gonzales v. O Centro Espírita Beneficente União do Vegetal, 546 U.S. 418, 430–32 (2006)) (calling for a "precise analysis" of asserted government interests).

194 As noted by amici, there are many school districts that could be used for this comparison. *See* Brief of the State of West Virginia, Commonwealth of Virginia, and 23 Other States as Amici Curiae in Support of Petitioners at 15–18, *Mahmoud*, 145 S. Ct. 2332 (No. 24-297).

negative outcomes in less restrictive states would also demonstrate necessity. When a state shows that granting a religious exemption would undermine the government's asserted interest, it suggests the more restrictive policy is needed to avoid bad outcomes.

This approach to negative outcomes reflects the Supreme Court's posture when evaluating necessity in the transition from *Sherbert* and *Yoder* to *Smith*. For example, in *United States v. Lee*,[195] the Court rejected an Amish employer's free exercise challenge to paying social security taxes. The Court reasoned that "[t]he tax system could not function" if religious objectors could obtain an exemption anytime "tax payments were spent in a manner that violates their religious belief."[196] Similarly, in *O'Lone v. Estate of Shabazz*,[197] the Supreme Court refused to grant religious accommodations to Muslim inmates. The Court upheld prison officials' policy against the requested accommodation because of the "adverse effects" on prison administration that would follow from allowing Muslim inmates an exemption to the prison rules at issue.[198] In both cases, the negative outcomes that would presumably follow from striking down the policies served to demonstrate their necessity.

In the context of *Mahmoud*, the school district could likewise have tried to meet its burden by showing how granting a religious exemption would undermine efforts to reduce bullying. The key to showing necessity in this form would be to demonstrate that districts without required instruction see higher rates of bullying against LGBTQ children. Under the hypothesis that required instruction is necessary, we would expect bullying rates to increase even if the required instruction were not sufficient to decrease bullying rates. Thus, the state could still have used DiD to perform this inquiry, assuming the appropriate data could be identified.[199]

B. *Catholic Charities Bureau v. Wisconsin Labor and Industrial Review Commission*

The second case is *Catholic Charities Bureau, Inc. v. Wisconsin Labor and Industrial Review Commission*.[200] In this case, Wisconsin

---

195 455 U.S. 252 (1982).

196 *Id.* at 260.

197 482 U.S. 342 (1987).

198 *Id.* at 352–53.

199 The Centers for Disease Control and Prevention collects data on bullying including toward LGBTQ youth, and district-level data is publicly available. *See Youth Risk Behavior Surveillance System: Data and Documentation*, CTRS. FOR DISEASE CONTROL & PREVENTION (Oct. 31, 2024), https://perma.cc/3HLG-S4VN.

200 145 S. Ct. 1583 (2025).

sought to deny Catholic Charities—the social ministry arm of the Catholic Church—an unemployment tax exemption.[201] The unemployment tax exemption is available to all organizations that operate "primarily for religious purposes."[202] Wisconsin determined that Catholic Charities does not operate primarily for religious reasons.[203] The Wisconsin Supreme Court agreed, reasoning that allowing religious motivations, rather than operations or activities, to be dispositive to a tax exemption determination "would cast too broad a net" and make the exemption unworkable.[204]

As described in an amicus brief from Ohio and eighteen other states, the Wisconsin Supreme Court's decision was grounded in "fear" that "religious claims will overwhelm the system if unchecked."[205] Thus, a fundamental question in the case was the necessity of a narrow construction of the tax exemption to the aims of the unemployment program. Using DiD, parties could have compared Wisconsin's program with unemployment tax exemptions in other states with similar language. For example, does Illinois' unemployment tax exemption—which takes a less restrictive interpretation[206]—require significantly higher taxes on other employers or otherwise damage the program's efficacy? Such evidence would have been valuable to judges in evaluating the necessity of Wisconsin's actions.

Just because DiD could have been applied in this case does not suggest how the inquiry would have gone. A particularly important factor to consider when dealing with religious exemptions is the unpredictable effect they can have: once a religious exemption is offered, it is hard to know how many people will take advantage of it. For this reason, an enduring argument against granting religious exemptions is that they will lead to a slippery slope of unending exemptions that swallow the rule.[207] What litigators and courts tend to overlook is that the veracity of this assumption

---

[201] Cath. Charities Bureau, Inc. v. Wis. Lab. & Indus. Rev. Comm'n, 3 N.W.3d 666, 673–74 (Wis. 2024), *rev'd*, 145 S. Ct. 1583 (2025).

[202] *Id.* at 676–80 (quoting WIS. STAT. § 108.02(15)(h) (2025)).

[203] *Id.* at 673–74.

[204] *Id.* at 679–82 (quotation marks omitted) (quoting Cath. Charities Bureau, Inc. v. Wis. Lab. & Indus. Rev. Comm'n, 987 N.W.2d 778, 792 (Wis. Ct. App. 2023)).

[205] Brief of Amici Curiae the State of Ohio and 18 Other States in Support of the Petitioners at 20, *Cath. Charities Bureau*, 145 S. Ct. 1583 (No. 24-154).

[206] *See* By the Hand Club for Kids, NFP, Inc. v. Dep't of Emp. Sec., 188 N.E.3d 1196, 1208 (Ill. App. Ct. 2020) (holding that a church's community ministry qualified for the state's religious-purposes unemployment tax exemption).

[207] *See Smith*, 494 U.S. at 888–90 (arguing that accepting a strict scrutiny framework as the default for free exercise challenges would be "courting anarchy").

largely depends on context. Cases in which the religious exemption is readily available or well-known (such as religious exemptions to vaccines) will produce more people taking advantage of it than obscure religious exemptions (such as tax exemptions for religious parsonages).[208] With this possible outcome in mind, crafting an appropriate DiD for Wisconsin would require an appropriate comparator state based on similarities between the prevalence of religious charities in the two states. If, for example, Illinois contained a comparable number of religious organizations registered as charities as did Wisconsin, Illinois would provide a useful baseline for comparison because tax exemptions would thus have a similar financial impact on the states.

Moreover, it is important to emphasize the role that qualitative evidence can play in supporting the assumptions required by DiD. Continuing the example above, a DiD comparing Wisconsin and Illinois could be further bolstered by presenting anecdotal evidence from employees at the Illinois Department of Employment Security. From the time that Illinois adopted a less restrictive interpretation, what did those employees see? Did the number of requested exemptions dramatically rise? This mixed-methods approach is valuable in social science generally[209] and would likewise bring more robust evidentiary justifications to the necessity inquiry.

C. *Oklahoma Statewide Charter School Board v. Drummond ex rel. Oklahoma*

The third and final case is *Oklahoma Statewide Charter School Board v. Drummond ex rel. Oklahoma*.[210] At issue in the case was an attempt to create the nation's first religious charter school. One of the key issues in the challenge was whether Oklahoma is justified in excluding the school from the state charter program.[211] The

---

208 In Illinois, an individual can obtain a religious exemption to vaccine requirements by filling out a two-page form signed by a physician. *See Illinois Certificate of Religious Exemption to Required Immunizations and/or Examinations Form,* ILL. DEP'T OF PUB. HEALTH, https://perma.cc/24J3-MM9V. By contrast, up to eighteen documents may be required to obtain the tax parsonage exemption. *See Exemptions for Religious Institutions*, COOK CNTY. ASSESSOR'S OFF., https://perma.cc/X46K-CJS2.

209 For a good example, see generally Barak-Corren & Berkman, *supra* note 115.

210 145 S. Ct. 1381 (2025) (mem.).

211 Petition for Writ of Certiorari at i, *Drummond*, 145 S. Ct. 1381 (mem.) (No. 24-396).

primary justification for Oklahoma's exclusion was an antiestablishment interest where Oklahoma must deny a charter to religious schools to comply with the federal Establishment Clause.[212]

This case squarely raises some limitations of DiD. Foremost is that a state's necessity argument cannot be evaluated against alternatives when no other state has adopted the counterfactual regime. In some contexts, perhaps using DiD to compare U.S. states with other countries may be a viable option given that some countries do allow religious charter schools.[213] The trouble in this case, however, is that the state interest being evaluated—avoiding violation of the federal Establishment Clause—is not readily comparable for any jurisdiction outside the United States. Given the unique history and tradition of the Establishment Clause,[214] only other U.S. states with a shared Establishment Clause heritage will be helpful in evaluating the necessity of Oklahoma's denial.

Another challenge to DiD in this case is the unanswered questions about Establishment Clause doctrine. The Supreme Court has identified coercion as "among the foremost hallmarks of religious establishments the framers sought to prohibit when they adopted the First Amendment."[215] Thus, if there is evidence that the religious character of charter schools can amount to coercion, it follows that Oklahoma's denial would be justified. But the Supreme Court has acknowledged that "[m]embers of th[e] Court have sometimes disagreed on what exactly qualifies as impermissible coercion in light of the original meaning of the Establishment Clause."[216] Thus, coercion has been a slippery term even for legal definition, making it particularly challenging to measure empirically.

In light of these limitations, DiD may largely be unhelpful in this context. Importantly, however, DiD's inapplicability does not free the state of its evidentiary burden. Instead, historical and anecdotal evidence may play a more prominent role in the state's argument. This other evidence illustrates a point worth emphasizing: the argument for increased employment of quantitative methods like DiD is not diametrically opposed to using history or

---

[212] *See* Drummond *ex rel.* State v. Okla. Statewide Virtual Charter Sch. Bd., 558 P.3d 1, 14–15 (Okla. 2024), *aff'd by an equally divided court*, 145 S. Ct. 1381 (2025) (mem.).

[213] *See, e.g.*, Shai Katzir & Lotem Perry-Hazan, *Promoting Politically Contested Change by Invisible Education Policies: The Case of Ultra-Orthodox Public Schools in Israel*, 50 OXFORD REV. EDUC. 658, 661–62 (2024) (discussing a religious public school system in Israel).

[214] For a primer, see generally Michael W. McConnell, *Establishment and Disestablishment at the Founding, Part I: Establishment of Religion*, 44 WM. & MARY L. REV. 2105 (2003).

[215] Kennedy v. Bremerton Sch. Dist., 142 S. Ct. 2407, 2428–29 (2022).

[216] *Id.* at 2429.

tradition as a method.[217] Instead, history and tradition are separate tools better fit for different constitutional puzzles. The state in a case like *Drummond* may find the historical record or traditional state practices more probative or revealing than trying to construct an illuminating DiD design. But the opposite is also frequently true: the historical record may establish very little about what the free exercise of religion means in a particular case. As already suggested, the government, in free exercise cases, gets to make a choice about how to justify its interests.[218] In many instances, DiD will provide an effective means for doing so.

CONCLUSION

Judges and scholars alike have found much to criticize about the appropriate method for deciding cases under the Free Exercise Clause. These critiques are especially sharp with regard to free exercise doctrine potentially (re)turning to strict scrutiny.[219] Within the free exercise case law, strict scrutiny has been about determining whether a government action is necessary to achieve its asserted aims. Although necessity has historically motivated constitutional reasoning in free exercise jurisprudence, relying on necessity does present a challenge. When government officials fail to bring tangible evidence that constrains the judicial inquiry, the necessity test may perversely invite subjective judicial discretion to fill in the gaps.

This Comment operationalizes necessity in free exercise through the empirical technique known as difference-in-differences. DiD leverages the U.S. federalist structure in a way that constrains judicial evaluation of necessity. Doing so benefits courts because it encourages litigants to offer empirical evidentiary justifications that allow for judicial discretion to be confined to evaluating facts instead of creating law. Adopting DiD to demonstrate necessity also benefits government actors, who are forced to consider the effectiveness of their desired policy before taking action. Finally, DiD likely benefits religious claimants,

---

[217] On this point, there seems to be agreement in the Supreme Court. *Compare* United States v. Rahimi, 144 S. Ct. 1889, 1921 n. 7 (2024) (Kavanaugh, J., concurring) (acknowledging "areas of constitutional law" where balancing tests rely on historical analysis), *with id*. at 1905 (Sotomayor, J., concurring) (acknowledging that "[h]istory has a role to play" in constitutional interpretation but rejecting a "rigid adherence to history" in such analyses), *and id*. at 1928 (Jackson, J., concurring) (same).

[218] *See supra* note 193 and accompanying text.

[219] *See Fulton*, 141 S. Ct. at 1882–83 (Barrett, J., concurring).

who can more readily expect that government officials will make their decisions with careful consideration of less restrictive alternatives used in other states.

The DiD technique is showcased using a current free exercise conflict over Ohio's law requiring flashing lights on Amish buggies. The results indicate that without its more restrictive buggy law, Ohio would have seen thirty-six more crashes per year on average—almost a 23% increase over the actual number of crashes. Given this substantive and statistically significant effect,[220] Ohio has a strong argument for the necessity of its buggy light law to advance its interest in Amish traffic safety.

Beyond the Amish controversy in Ohio, DiD also looks promising for evaluating necessity in other free exercise cases, including some recently decided by the Supreme Court. Applying DiD to these conflicts demonstrates how DiD can be used by litigants even when expressive or unquantifiable interests are at stake. Moreover, the use of DiD need not detract from existing methods. When the history or tradition say little to address necessity in a free exercise controversy, DiD may provide a helpful and complementary means to resolve the case.

While the integration of DiD into the necessity inquiry offers solutions, it also raises new questions for scholars to continue exploring. For example, how might this mode of analysis work for federal officials who may not have a good comparison at the state level? Beyond the Free Exercise Clause, how might DiD work within other constitutional contexts that rely on heightened scrutiny, such as equal protection or substantive due process? The fact that DiD raises additional questions is both motivating and assuring. It is motivating because there is much work left to be done in generating more and better evidence in lawmaking. But it should also be assuring that in offering one answer to a constitutional question, we invite additional discussion on the new questions that answer may raise. Indeed, the Constitution's broadly worded provisions often encourage ongoing debates. Given the importance of the free exercise of religion, we should be eager to seek out the best evidence to inform those debates.

---

[220] *See supra* note 176 and accompanying text.

APPENDIX

TABLE A: FULL MONTH-LEVEL SUMMARY STATISTICS OF TOTAL CRASHES INVOLVING BUGGIES BY STATE

Time	State	Total Crashes Involving Buggies
Oct. 2021	Ohio	11
Nov. 2021	Ohio	12
Dec. 2021	Ohio	16
Jan. 2022	Ohio	12
Feb. 2022	Ohio	10
Mar. 2022	Ohio	8
Apr. 2022	Ohio	10
May 2022	Ohio	11
June 2022	Ohio	10
July 2022	Ohio	8
Aug. 2022	Ohio	13
Sept. 2022	Ohio	11
Oct. 2022	Ohio	10
Nov. 2022	Ohio	14
Dec. 2022	Ohio	11
Jan. 2023	Ohio	10
Feb. 2023	Ohio	4
Mar. 2023	Ohio	7
Apr. 2023	Ohio	5
May 2023	Ohio	9
June 2023	Ohio	7
July 2023	Ohio	5
Aug. 2023	Ohio	10
Sept. 2023	Ohio	12
Oct. 2021	Michigan	4
Nov. 2021	Michigan	5
Dec. 2021	Michigan	5
Jan. 2022	Michigan	0
Feb. 2022	Michigan	1
Mar. 2022	Michigan	1
Apr. 2022	Michigan	2
May 2022	Michigan	1
June 2022	Michigan	2
July 2022	Michigan	1
Aug. 2022	Michigan	3

Sept. 2022	Michigan	3
Oct. 2022	Michigan	4
Nov. 2022	Michigan	2
Dec. 2022	Michigan	3
Jan. 2023	Michigan	1
Feb. 2023	Michigan	3
Mar. 2023	Michigan	4
Apr. 2023	Michigan	6
May 2023	Michigan	2
June 2023	Michigan	1
July 2023	Michigan	4
Aug. 2023	Michigan	4
Sept. 2023	Michigan	5
Oct. 2021	Kentucky	0
Nov. 2021	Kentucky	2
Dec. 2021	Kentucky	0
Jan. 2022	Kentucky	1
Feb. 2022	Kentucky	4
Mar. 2022	Kentucky	2
Apr. 2022	Kentucky	3
May 2022	Kentucky	5
June 2022	Kentucky	2
July 2022	Kentucky	1
Aug. 2022	Kentucky	1
Sept. 2022	Kentucky	2
Oct. 2022	Kentucky	4
Nov. 2022	Kentucky	2
Dec. 2022	Kentucky	2
Jan. 2023	Kentucky	2
Feb. 2023	Kentucky	2
Mar. 2023	Kentucky	0
Apr. 2023	Kentucky	0
May 2023	Kentucky	1
June 2023	Kentucky	2
July 2023	Kentucky	1
Aug. 2023	Kentucky	1
Sept. 2023	Kentucky	3

TABLE B: OHIO BUGGY LAW DID REGRESSION

	Average Monthly Crashes
Ohio x Post-Law	−2.583* (−2.57)
Ohio	8.875*** (12.48)
Post-Law	0.250 (0.43)
Total Observations	72
Mean Crashes (Ohio)	11

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$