NEW YORK UNIVERSITY
LAW REVIEW

VOLUME 99 JUNE 2024 NUMBER 3

ARTICLES

CONSTITUTIONAL CONSEQUENCES

NETTA BARAK-CORREN[†] & TAMIR BERKMAN[‡]

For over two hundred years of Supreme Court doctrine, judges and scholars have tried to figure out how the Court's rulings impact ordinary citizens. Yet the answers often seem to depend on whose opinion or even which press releases you read. How can we actually measure the consequences of constitutional decisions?

This Article provides a new methodological inroad to this thicket—one which triangulates a nationwide field experiment, a longitudinal public opinion survey, and litigation-outcome analysis. We do so while focusing on a recent set of developments at the intersection of religious freedom and anti-discrimination law that transpired in Fulton v. City of Philadelphia *(2021).*

We find that Supreme Court decisions can have substantial behavioral and legal effects beyond a seemingly narrow holding. In Fulton*, the Court avoided deciding the equality-religion conflict at the heart of the case for a fact-specific decision that should have been easy to circumvent. Yet our results suggest that the Court's audience focused on the bottom-line message of the decision rather than the holding. Across the nation, foster care agencies became less responsive to same-sex couples. The*

[†] Professor of Law and member of the Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, and Law, Ethics, and Public Policy Fellow at Princeton University Center for Human Values.

[‡] Hebrew University Law School alumnus and Associate at Meitar. For helpful comments and suggestions, we thank Danielle Allen, Benjamin Edelson, Donald Green, David Hoffman, Yael Kariv-Teitelbaum, Hajin Kim, Jonathan Mazur, Roni Porat, Zalman Rothschild, Sonja Starr, Tess Wilkinson-Ryan, and workshop participants in Berkeley, Chicago, Harvard, Michigan, Penn, Princeton, the 2022 and 2023 Conferences on Empirical Legal Studies, the 2023 Nootbaar Institute Religious Freedom Conference, and the 2023 Law and Religion Roundtable. Chen Agassi, Yonah Becker, Eli Benin, Ziv Fox, David Freud, Elliot Friedman, Neomi Goldberg, Shira Laurence, Caroline Ribet, Noam Shlomai-Kunitz, Shiraz Sultan, Tal Zandman, and Jonathan Zisk provided excellent research assistance.

public became more supportive of religious service refusals. And courts and litigants resolved all open disputes between equality-seeking governments and refusing religious agencies in favor of the agencies.

Our findings contribute to the development of an empirical approach to constitutional doctrine. Constitutional questions often require determining whether the harm to, or burden on, an individual or group is justified by a compelling state interest—and whether the means are narrowly tailored to that end. These tests often hinge on evidence, yet the Court rarely offers parties guidelines for substantiating their interests at the right level of precision. Our work provides both data and empirical tools that inform the application of this test in the realm of free exercise doctrine, equality law, and beyond.

INTRODUCTION 787
I. THE LAY OF THE LAND 796
II. THE FIELD EXPERIMENT 803
A. *An Introduction to Field Experiments in Law, and Our Design* 804
1. *The Sample* 807
2. *Procedure* 808
3. *Coding the Data* 810
B. *Results* 811
1. *Baseline Discrimination* 811
2. *Did the* Fulton *Decision Increase Discrimination Towards Same-Sex Couples?* 814
a. The Effect of *Fulton* on Antidiscrimination Only Jurisdictions 815
b. Changes Among the Three Other Legal Regimes 816
3. *Converging Evidence from a Longitudinal Survey* 817
C. *Discussion* 820
III. EVIDENCE FROM LITIGATION: HOW PARTIES THOUGHT *FULTON* APPLIED 822
IV. CONSEQUENTIAL CONSTITUTIONAL LAW 826
A. *Converging Evidence on Social Consequences of the Law* 827
1. *Effects on Service Providers* 828
2. *Effects on LGBTQ Individuals* 829
B. *The Interplay Between Social Consequences and Legitimacy Consequences* 831
1. *The Consequential Challenges for Preserving Legitimacy* 832

2. *The Social Price of Preserving Legitimacy* 834
C. *Incorporating Consequences into Constitutional Doctrine* 836
1. *The Problematic Application of the Compelling State Interest Test* 838
2. *Concretizing Compelling Interests Empirically* . . . 842
3. *Testing the Means to Ensure Narrow Tailoring* . . . 844
D. *Limitations of Consequential Constitutionalism* 846
CONCLUSION 848
APPENDIX 849

INTRODUCTION

"It is the lack of an empirical footing that is and always has been the Achilles heel of constitutional law."

—Richard Posner, 1998[1]

For over two hundred years of Supreme Court doctrine, judges and scholars have tried to figure out how the Court's rulings impact ordinary citizens.[2] Even when a decision seems to have straightforward implications—this person gets relief, or the government cannot make this law—the downstream consequences are hard to predict and harder to measure.[3] The Justices try to shape the behavior of their stakeholders: Congress should legislate, or not;[4] prosecutors should disclose, or

[1] Richard A. Posner, *Against Constitutional Theory*, 73 N.Y.U. L. REV. 1, 21 (1998).

[2] *See, e.g.*, Aaron Tang, *Consequences and the Supreme Court*, 117 NW. U. L. REV. 971, 1018–30 (2023) (discussing the use, lack of, and merits of consequentialist reasoning in Supreme Court opinions). *See generally* Mark S. Kende, *Constitutional Pragmatism, the Supreme Court, and Democratic Revolution*, 89 DENV. U. L. REV. 635, 635–70 (2012) (detailing different pragmatic approaches to legal reasoning).

[3] *See, e.g.*, Albert W. Alschuler, *Plea Bargaining and Its History*, 79 COLUM. L. REV. 1, 37–40 (1979) (arguing that the Supreme Court's "due process revolution" that strengthened rights for criminal defendants perversely contributed to the rise of plea bargaining, a process in which defendants receive fewer protections, because prosecutors spent more time on pre-trial motions, appeals, and post-conviction proceedings); Roseanna Sommers & Vanessa K. Bohns, *The Voluntariness of Voluntary Consent: Consent Searches and the Psychology of Compliance*, 128 YALE L.J. 1962, 2011–19 (2019) (finding that the behavioral assumptions underpinning search and seizure doctrine are erroneous as a matter of empirical evidence).

[4] *See Table of Laws Held Unconstitutional in Whole or in Part by the Supreme Court*, CONST. ANNOTATED, https://constitution.congress.gov/resources/unconstitutional-laws [https://perma.cc/DQK6-B8FR] (detailing every law, federal, state, or local, that the Supreme Court has ever declared unconstitutional); Roger Clegg, *Introduction: A Brief Legislative History of the Civil Rights Act of 1991*, 54 LA. L. REV. 1459, 1459–63 (1994) (describing a series of Supreme Court decisions that interpreted statutes narrowly and precipitated Congress's passing of the Civil Rights Act of 1991).

withhold;[5] agencies should explain themselves;[6] lower courts should conform;[7] and the people should be more or less decorous, depending on the context.[8] But however apparently clear the Court's instructions, the execution is inevitably muddier.

Discrimination rulings stand out as especially challenging in this respect, because of the sheer volume and complexity of the social manifestations of discrimination. It is one thing for the Court to announce that school segregation is unconstitutional, and another thing to try to remove discrimination from American public schools.[9] It is one thing to declare "colorblindness" and forbid the use of race in university admissions decisions, and another thing to ensure equal opportunities and outcomes regardless of race.[10] Similarly, to

---

[5] *See, e.g.*, Brady v. Maryland, 373 U.S. 83, 87 (1963) (holding that prosecutors are required to turn over evidence that is favorable to the accused and material to guilt or punishment); Bruton v. United States, 391 U.S. 123, 137 (1968) (ruling that the prosecution's introduction of a non-testifying co-defendant's confession violated the defendant's confrontation clause rights).

[6] *See, e.g.*, Motor Vehicle Mfrs. Ass'n v. State Farm Mutual Auto. Ins. Co., 463 U.S. 29, 48 (1983) ("We have frequently reiterated that an agency must cogently explain why it has exercised its discretion in a given manner . . . ."); Dep't of Homeland Sec. v. Regents of the Univ. of Cal., 140 S. Ct. 1891, 1916 (2020) (remanding because an agency failed to comply "with the procedural requirement that it provide a reasoned explanation for its action" and giving it an opportunity to "consider the problem anew").

[7] *See, e.g.*, Sarah Herman Peck, Cong. Rsch. Serv., R44618, Post-*Heller* Second Amendment Jurisprudence 12–13, 17 (2019) (noting that *District of Columbia v. Heller*, 544 U.S. 570 (2008), a landmark Second Amendment case, left critical issues unresolved and lower courts have taken various approaches to applying its reasoning); United States v. Chester, 628 F.3d 673, 688–89 (4th Cir. 2010) (Davis, J., concurring) ("*Heller* has left in its wake a morass of conflicting lower court opinions regarding the proper analysis to apply to challenged firearms regulations.").

[8] *See, e.g.*, Cohen v. California, 403 U.S. 15, 16, 26 (1971) (holding that a statute criminalizing the display of "offensive conduct" violated the freedom of speech); Chaplinsky v. New Hampshire, 315 U.S. 568, 573–74 (1942) (upholding an arrest for the use of "fighting words" because they are "likely to cause a breach of the peace").

[9] *See* Charles J. Ogletree, Jr., *All Deliberate Speed: Reflections on the First Half-Century of* Brown vs. Board of Education, 66 Mont. L. Rev. 283, 289–92 (2005) (explaining that the "underlined hypocrisy" of *Brown* was that it mandated that desegregation move slowly and detailing the enforcement challenges of desegregating education); John H. Blume, Sheri Lynn Johnson & Ross Feldmann, *Education and Interrogation: Comparing* Brown *and* Miranda, 90 Cornell L. Rev. 321, 338 (2005) (explaining that the *Brown* decision "allowed states to easily circumvent goals that the Court had in mind").

[10] *Compare* Students for Fair Admissions, Inc. v. President and Fellows of Harvard Coll., 143 S. Ct. 2141, 2175, 2277 (2023) (striking down affirmative action in college admissions because upholding such programs would create "a judiciary that picks winners and losers based on the color of their skin"), *with* Devon W. Carbado & Cheryl I. Harris, *The New Racial Preferences*, 96 Calif. L. Rev. 1139, 1168 (2008) (arguing that Michigan and California propositions aimed at banning the use of race admissions end up "conferring a preference for applicants for whom race does not matter"), *and* Kimberly Jenkins Robinson, *The Constitutional Future of Race-Neutral Efforts to Achieve Diversity and Avoid Racial Isolation in Elementary and Secondary Schools*, 50 B.C. L. Rev. 277, 309 (2009) ("[T]he adoption of a

declare that same-sex relationships deserve equal treatment under the law does not mean that every citizen gives up their prejudice.[11] Understanding cause and effect in a complex society, especially when the issues potentially affect the daily choices of millions of people, is a longstanding methodological challenge—and yet, that understanding is also crucial. Judges, advocates, and ordinary people care a lot about what a constitutionalized discrimination ruling will actually mean for American society—who will benefit, who will suffer, what will change, and what will endure.

This Article offers an unusual and novel methodological inroad to this thicket through a combination of a field experiment, a longitudinal public opinion survey, and a complete analysis of litigation outcomes, while focusing on a recent set of developments that have unfolded at the intersection of religious freedom and antidiscrimination law.

The Supreme Court has been recently asked to consider a set of claims from religious parties that would permit them to refuse service to LGBTQ clients on the basis of their religious opposition to same-sex relationships.[12] It has repeatedly written decisions that on their face seem narrowly tailored to the questions presented,[13] but which are sometimes received by advocates as portending dramatic changes on the ground.[14]

---

colorblind approach to the Constitution today would ignore the structural inequalities that converge in racially isolated schools to create inferior educational opportunities."), *and* Amy Stuart Wells & Erica Frankenberg, *The Public Schools and the Challenge of the Supreme Court's Integration Decision*, 89 Phi Delta Kappan 178, 183 (2007) (noting that Justices advocating for colorblindness in an affirmative action case ignore the "centrality of structural inequality and its relationship to racial segregation").

[11] *See infra* Section II.B (discussing the results from our field experiment which show pervasive discrimination against same-sex couples among foster care agencies).

[12] *See infra* Part I.

[13] *See, e.g.*, Thomas G. Donnelly, *Supreme Court Legitimacy: A Turn to Constitutional Practice*, 47 BYU L. Rev. 1487, 1507–17 (2022) (discussing Justice Roberts's approach to coalition building, including narrow judicial "incrementalism"); Jeannie Suk Gersen, *The Supreme Court's Surprising Term*, New Yorker (June 27, 2021), https://www.newyorker.com/magazine/2021/07/05/the-supreme-courts-surprising-term [https://perma.cc/JYN6-7SE8] (explaining that, to preserve legitimacy and avoid partisanship, the Justices "reach[ed] broad agreement on narrow issues").

[14] *See, e.g.*, Louise Melling, *The New Faith-Based Discrimination*, Bos. Rev. (Dec. 14. 2022), https://www.bostonreview.net/articles/the-new-faith-based-discrimination [https://perma.cc/H9QY-9CWG] (arguing, from the perspective of a civil liberties advocate, that the decisions in *Fulton* and several other recent religious liberty rulings have "grave" implications "for our civil rights laws and those they aim to protect" and undermine precedent); *Unanimous Supreme Court Protects Foster Moms & 200-Year-Old Ministry*, Becket (June 17, 2021), https://www.becketlaw.org/media/unanimous-supreme-court-protects-foster-moms-200-year-old-ministry [https://perma.cc/5FPK-LEAM] (noting, from the perspective of the organization that represented CSS, that *Fulton* "is a strong message in favor of religious freedom").

Fulton v. City of Philadelphia is one such seemingly incremental, but potentially transformative, case.[15]

Fulton's story began when Philadelphia terminated the contract of a Catholic foster care agency that refused to serve same-sex couples.[16] The City argued that this policy violated its antidiscrimination laws; the Catholic agency responded that it was the City which was violating its free exercise rights.[17] Ultimately, the Supreme Court unanimously agreed with the agency, ruling that the agency should have been granted a religious exemption from the antidiscrimination policy.[18] Writing for the Court, Chief Justice Roberts focused on specific contractual language rather than the conflict between antidiscrimination and free exercise.[19]

The *Fulton* opinion limited its holding to the facts of the case, avoiding the conflict at the heart of the dispute so much that the more conservative Justices were openly frustrated. As Justice Alito quipped in his concurrence, "[t]his decision might as well be written on the dissolving paper sold in magic shops."[20] Justice Gorsuch explained that "with a flick of a pen, municipal lawyers may rewrite the City's contract to close the . . . loophole."[21] And yet, even though *Fulton* may seem to some like a narrowly decided case, what is limited and what is broad turns on where one starts.[22] Advocates for gay couples worried that *Fulton* would deliver a setback in the struggle for LGBTQ equality;[23]

---

15 *See* Fulton v. City of Philadelphia, 141 S. Ct. 1868 (2021).

16 *Id.* at 1874.

17 *Id.* at 1875–76.

18 *Id.* at 1882.

19 *See id.* at 1881.

20 *Id.* at 1887 (Alito, J., concurring).

21 *Id.* at 1930 (Gorsuch, J., concurring).

22 This view of the Court's opinion was voiced first and foremost by Justices Alito and Gorsuch, who concurred with the outcome but disagreed with the reasoning. *See infra* notes 130–31. Some civil rights organizations shared a similar reaction. *See, e.g.*, Christopher Vasquez, *NCLR Relieved by Narrow SCOTUS Ruling in Fulton Allowing Governments to Prohibit Anti-LGBTQ Discrimination*, Nat'l Ctr. Lesbian Rts., https://www.nclrights.org/about-us/press-release/nclr-relieved-by-narrow-scotus-ruling-in-fulton-allowing-governments-to-prohibit-anti-lgbtq-discrimination [https://perma.cc/A84U-NLNS] (celebrating that the narrow ruling would not lead to additional exemptions); James Esseks, *Supreme Court Again Rejects a License to Discriminate*, ACLU (June 17, 2021), https://www.aclu.org/news/lgbtq-rights/supreme-court-again-rejects-a-license-to-discriminate [https://perma.cc/UX5V-M4HH] (celebrating the narrow ruling similarly, noting that "governments can . . . continue to enforce their non-discrimination laws in all contexts").

23 *See, e.g.*, Adam Liptak, *Supreme Court Backs Catholic Agency in Case on Gay Rights and Foster Care*, N.Y. Times (June 17, 2021), https://www.nytimes.com/2021/06/17/us/supreme-court-gay-rights-foster-care.html [https://perma.cc/4JAJ-YBXX] ("The decision . . . was a setback for gay rights and further evidence that religious groups almost always prevail in the current court."); Elie Mystal, *The Supreme Court Strikes Another Blow to the Separation*

and those for religious groups celebrated what they saw as an increasingly permissive Court.[24]

Whether *Fulton* is revolutionary and consequential or not thus may depend on whose opinion you read, or even which press releases you read.[25] You might be tempted to throw up your hands.

We do not think that the question of what *Fulton* and these other cases do is all that intractable, nor that constitutional law should be indifferent to its consequences simply because measuring them is hard. In this Article, we evaluate this question in three steps. First, we conducted a nationwide field experiment measuring the actual effects of *Fulton*'s ruling on foster care agencies and their treatment of same-sex and opposite-sex couples. We triangulated the results of the field experiment with the results from a longitudinal survey experiment of the American public that we conducted in parallel to the field experiment, and yielded the same pattern of results. Finally, we examined how similar disputes between governments and religious foster care agencies were resolved in *Fulton*'s wake.

We first conducted a field experiment to measure the effects of *Fulton*. To do so, we contacted all 1,905 foster care agencies in the United States with outreach emails purporting to be from same-sex and opposite-sex couples interested in fostering a child. We initially did that before *Fulton* was decided, but after its oral argument. We then contacted the agencies after the decision, again purporting to be either a same-sex or an opposite-sex couple. Then, we observed what the agencies did. Our basic questions: How do agencies in different legal jurisdictions behave toward same-sex, as opposed to opposite-sex couples? Did the agencies change how they responded to inquiries from before to after *Fulton*?

This is what we found: First, even before *Fulton*, foster care agencies across the nation were less likely to respond to same-sex couples than

of Church and State, Nation (June 22, 2022), https://www.thenation.com/article/society/supreme-court-religious-schools [https://perma.cc/EFU2-SVSH].

[24] *See, e.g.*, Ethan Tong & Jorge Gomez, *Religious Freedom in Decline? Major Wins at the Supreme Court Suggest the Tide Is Turning*, First Liberty (Aug. 12, 2022), https://firstliberty.org/news/major-wins-suggest-the-tide-is-turning [https://perma.cc/H4JX-DE3K] (describing *Fulton* as a "crucial religious freedom victor[y]"); Emilie Kao, *Supreme Court Decision Is a Win for Religious Freedom*, Heritage Found. (June 23, 2021), https://www.heritage.org/courts/commentary/supreme-court-decision-win-religious-freedom [https://perma.cc/QY3B-QAS4].

[25] In Part I, *infra*, we discuss the additional Supreme Court rulings in the religion-LGBTQ rights space, *Masterpiece Cakeshop, Ltd. v. Colorado Civil Rights Commission*, 138 S. Ct. 1719 (2018), and *303 Creative v. Elenis*, 143 S. Ct. 2298 (2023), which share this feature as well.

they were to opposite-sex couples.[26] The same couple approaching an agency was 1.12 times more likely to get a response if they were straight than if they were not. On average, foster care agencies are, as advocates have long argued,[27] less responsive to gay couples. But the effect varies greatly by jurisdiction. Agencies operate in states that differ in how they balance questions of antidiscrimination and religious freedom in foster care. Some agencies are bound by rules mandating the equal treatment of prospective LGBTQ foster parents, while other agencies operate in jurisdictions that never enacted such rules. Some jurisdictions exempted religious agencies from having to comply with antidiscrimination laws; others did not.

After laboriously coding those rules, we found some unsurprising results. Prior to *Fulton*, in states like Minnesota that prohibited discrimination and did not offer religious exemptions, there were small differences in response rates—around 3%.[28] But in states like Michigan (which enacted a religious exemption law) and Texas (with no antidiscrimination law at all), we found material differences in response rates—18% and 12% respectively.[29] These differences persist even when controlling for various socio-demographic and local attitudes towards gay rights that naturally vary between these different jurisdictions.[30]

As we will discuss in more detail, inferring causality is a tricky business, especially because the dynamics between local attitudes and local law shape behavior in ways that are hard to measure. But our design enables us to look at change associated with *Fulton* itself: Did

---

[26] *See infra* Section II.B.1. Data on states' regulations of foster care agencies in the context of antidiscrimination and religious exemptions was collected in an earlier study. *See* 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, 1013–16 (2022).

[27] *See, e.g.*, Cheryl Corley, *Illinois, Catholic Agencies at Odds Over Gay Adoptions*, NPR (July 5, 2011), https://www.npr.org/2011/07/05/137622143/illinois-catholic-agencies-at-odds-over-gay-adoptions [https://perma.cc/7RA8-C52Q] (noting that several Catholic adoption services in Illinois only placed children with straight individuals); Fla. Dep't of Child. & Fams. v. Adoption of X.X.G., 45 So. 3d 79, 92 (Fla. Dist. Ct. App. 2010) (finding that a law that expressly banned gay adoptions violated the state's constitution); Abbie E. Goldberg, Jordan B. Downing & Hannah B. Richardson, *The Transition from Infertility to Adoption: Perceptions of Lesbian and Heterosexual Couples*, 26 J. Soc. & Pers. Relationships 938, 958 (2009) (explaining that queer women faced challenges finding "gay-friendly adoption agenc[ies]"); Lori A. Kinkler & Abbie E. Goldberg, *Working with What We've Got: Perceptions of Barriers and Supports Among Small-Metropolitan-Area Same-Sex Adopting Couples*, 60 Fam. Rels. 387, 392–93 (2011) (describing the difficulty of finding an agency willing to work with a same-sex couple).

[28] *See infra* Section II.B.1.

[29] *See infra* Section II.B.1.

[30] *See infra* note 119.

Fulton change the behavior of foster care agencies that were previously prohibited from discriminating? We found that after *Fulton*, the gap in positive responses between same-sex and opposite-sex couples tripled in jurisdictions like Minnesota, which previously did not offer religious exemptions from antidiscrimination law—from 3% to 9% on average.[31] In states that already had laws protecting religious refusal, like Texas, the average gap widened from 12% to 20%.[32] In states that had no pertinent antidiscrimination laws, like Ohio, the gap increased slightly from 12% to 14%.[33] In the small cohort of agencies in states with both antidiscrimination and religious refusal laws (nearly all of them in Michigan), disparities narrowed after *Fulton* from 18% to 2%.[34] Except for this last result, we find corresponding trends in the longitudinal public opinion survey we conducted. The survey—our second layer of evidence—shows growing public support for religious service refusal post-*Fulton* in all jurisdictions.[35]

Hence, *Fulton* appears to have mattered a lot—shifting the behavior of foster care agencies in states whose legal rules were affected the most. But it *also* mattered in jurisdictions that were not supposed to be influenced by the decision: where antidiscrimination rules were never enacted or already included exemptions. In a sense, partisans on all sides were right: The Court's decisions, even when framed narrowly, affect behavior in the world, beyond that of the stakeholders who it legally bound. But the pattern of those effects is more complicated than the common slogans.

Our third layer of evidence reveals that the experimental and survey findings fit with the behavior of parties who litigated similar cases in the aftermath of *Fulton*. First, Philadelphia and Catholic Social Services (CSS) settled after the *Fulton* decision,[36] and Philadelphia paid CSS two million dollars and renewed its contract without requiring that the agency comply with the city's nondiscrimination policy.[37]

---

31 *See infra* Section II.B.2.

32 *See infra* Section II.B.2.

33 *See infra* Section II.B.2.

34 *See infra* Section II.B.2.

35 *See infra* Section III.A.2.

36 Joint Motion for Consent Judgment at 2, Fulton v. City of Philadelphia, 320 F. Supp. 3d 661 (E.D. Pa. 2021) (No. 18-2075), *rev'd*, 141 S. Ct. 1868 (2021), https://www.aclu.org/wp-content/uploads/legal-documents/077_joint_motion_for_consent_judgement.pdf [https://perma.cc/RT92-KTU9].

37 *Id.* at 4–5; Victoria A. Brownworth, *Philly Agrees to Pay $2M in Settlement Over LGBTQ Foster Care Lawsuit*, Pa. Cap. Star (Nov. 24, 2021), https://www.penncapital-star.com/civil-rights-social-justice/philly-agrees-to-pay-2m-in-settlement-over-lgbtq-foster-care-lawsuit [https://perma.cc/J4P5-4Y7G].

"[T]he city chose not to continue the fight"[38] despite the fact that according to Justice Alito's concurrence, it could have just removed its never-used exemption clause from the contract to prohibit the agency from discriminating.[39] That the parties chose to settle suggests that they may have viewed the decision as a more decisive position on the conflict.

Furthermore, this outcome was not unique. Our analysis of litigation outcomes indicates that *all* disputes between equality-seeking governments and refusing religious agencies were resolved after *Fulton* in the shadow of a presumption in favor of religious exemptions, even if the circumstances had nothing to do with the holding. In conversations we conducted with lawyers who represented parties in these cases, we learned that stakeholders understood *Fulton* as a clear signal from the Court: Governments must exempt refusing agencies from antidiscrimination rules.[40]

Our three layers of evidence complement each other: We now have both data from litigation and data from the entire field of foster care agencies (most of which would have never reached the courtroom)—as well as from public opinion. These data reveal both short and longer-term consequences in pertinent jurisdictions—states like Minnesota, where *Fulton* modified the law in practice—as well as in seemingly irrelevant jurisdictions, like Texas, where the decision did not alter the state of the law but nonetheless influenced behavior. Furthermore, our work converges with previous evidence on the effects of Supreme Court decisions and religious exemption laws, and as such it replicates and extends previous findings.[41]

Zooming out from the particular doctrinal application, this Article contributes to the development of a novel constitutional methodology, deeply rooted in rigorous methods of causal inference. Black letter constitutional law often asks if a governmental burden on an individual or group can be justified by a compelling state interest.[42] However, the

---

38 Brownworth, *supra* note 37.

39 *Fulton*, 141 S. Ct. at 1887 (Alito, J., concurring).

40 *See infra* note 150.

41 *See infra* Section IV.A.1.

42 In free exercise cases, laws that restrict the exercise of religion and are not generally applicable are subject to strict scrutiny, which requires governments to show that they have compelling interests in burdening the exercise of religion. *See generally* Emp. Div., Dep't of Hum. Res. v. Smith, 494 U.S. 872, 878, 882 (1990) (holding that generally applicable laws need not overcome strict scrutiny). Although it is not clear exactly what the government needs to show to meet its burden, case law suggests that the state needs to provide evidence of its compelling interest. *See* Sherbert v. Verner, 374 U.S. 398, 407 (1963) (holding that the state must present more than "a possibility" of harm to justify a restriction on the exercise of religion); Wisconsin v. Yoder, 406 U.S. 205, 221 (1972) (holding that in free exercise cases,

Court rarely offers parties guidelines for substantiating their interests at the right level of precision.[43] Our work paves a path forward by providing both data and empirical tools that inform the application of this test.

In the context of religion-equality conflicts, our data support a general conclusion that governments *often* have a compelling interest not to offer religious exemptions from antidiscrimination laws, as these exemptions are, as we find, associated with increased discrimination. But our finding is more subtle, and perhaps destabilizing of current doctrine.[44] The strength of states' interests—at least those that rely on magnitude of discrimination and access to services—could vary given the differences between them, including the number of operating agencies (ranging from a handful to more than a hundred agencies per state), the share of religious agencies (ranging from none to 100% of agencies per state), and the magnitude of the negative disparities between same-sex and opposite-sex couples (ranging, according to our estimates, from none to 30%). In other words, our results suggest that the compelling interest test requires a much more detailed, careful, and evidence-based analysis than that typically conducted by states and courts.

None of our findings imply that narrowly framed decisions have *no* significance for law and policy, or that these decisions will *always* be interpreted broadly by the Court's multiple audiences. We suggest that our results are particularly relevant given several contextual dimensions that were present in *Fulton* and could emerge again. First, the decision was persuasively portrayed in relevant circles as a big victory.[45] This is arguably easier when the holding is ambiguous, vague, or insincere (*Fulton* was explicitly criticized as such by scholars and lawyers). Second, where a narrowly framed decision comes as part of a series of decisions whose outcomes build up a consistent pattern, such as the series of religion-supportive cases that the Supreme Court handed down in recent years, stakeholders can connect the dots and pay more attention to the outcome than

---

the court cannot "accept . . . a sweeping claim" by the state but instead must "searchingly examine" the state's interest). Slippery slope justifications generally do not satisfy the government's burden. *See* Gonzalez v. O Centro Espiritia Beneficente Uniao do Vegetal, 546 U.S. 418, 436 (2006) (rejecting the government's argument that "[i]f I make an exception for you, I'll have to make one for everybody, so no exceptions"). For a detailed discussion of the compelling interest test in free exercise cases, see *infra* Section IV.C.1.

[43] *See infra* Section IV.C.1 (detailing the Court's failure to articulate guidelines for substantiating compelling governmental interests).

[44] *See infra* Sections IV.C.2–.3.

[45] *See supra* notes 22–24.

the reasoning. In this present moment of free exercise doctrine, the Court appears to signal a broader agenda than the reasoning offered by each decision alone.

Making constitutional doctrine turn on real, rather than hypothetical, consequences is not an exercise for the faint-hearted. Though constitutional law has often been framed with reference to empirical truths—think, the separate-and-unequal bases of *Brown*[46]—it has rarely checked its aspirations against its outcomes. Consequential constitutionalism might result in outcomes that advocates—for equality, liberty, order, expression, anti-racism—will find distasteful. We need to carefully consider whether (or when) consequences really are what we should care about when evaluating constitutional rules. Perhaps parts of constitutional law ought to distance themselves from consequential concerns for normative reasons or due to practical limitations. Or perhaps not: A consequential constitutionalism might help to ground courts better in the real world. We consider these objections and concerns as a part of making a case for taking consequences more seriously, in civil rights and liberties doctrines and more broadly.

The remainder of this Article proceeds in four parts. Part I gives an overview of the legal background and explores the different constraints foster care agencies navigate in different jurisdictions. Part II describes our experiment's method and results in detail. In Part III we trace the litigation that followed *Fulton* and describe its outcomes. In Part IV we discuss the converging evidence and situate our findings in the context of prior empirical work, discuss the theoretical and normative implications of our work for the Supreme Court, and explain how our work contributes to developing and applying the Court's free exercise doctrine.

I
The Lay of the Land

The City of Philadelphia contracts with private agencies to help place children with foster families.[47] These agencies are paid with taxpayer funds to perform this public role.[48] The contracts explicitly

---

[46] *See* Brown v. Bd. of Educ., 347 U.S. 483, 495 (1954) ("Separate educational facilities are inherently unequal.").

[47] *See Foster Care Licensing Agencies*, Phila. Dep't Hum. Servs., https://www.phila.gov/media/20220915154821/DHS_Philadelphia_Foster_Care_Agencies_091422.pdf [https://perma.cc/K7BL-46EY].

[48] Brief for the Respondent at 1, Fulton v. City of Philadelphia, 141 S. Ct. 1868 (2021) (No. 19-123).

forbid discrimination on the basis of protected characteristics, including sexual orientation.[49]

Fulton came to federal court in 2018 when Philadelphia cut ties with two religious foster care agencies after learning that they would not work with LGBTQ couples.[50] Indeed, the City's desire to prevent discrimination interfered with the agencies' freedom to abide by their religious beliefs. A few months later, one of the agencies and three foster parents sued for injunctive relief and damages, arguing that Philadelphia had violated the Free Exercise Clause, the Free Speech Clause, and Pennsylvania's religious freedom statute.[51]

Both the district court and the circuit court ruled in favor of the City. These lower court decisions were controlled by *Employment Division v. Smith*, which held that neutral laws of general applicability cannot be challenged under the Free Exercise Clause without a showing of religious animus.[52] Laws that are not generally applicable need to pass strict scrutiny, where the burden is on the government to show that the law is narrowly tailored to a compelling state interest in order to survive.[53] The district court concluded that the City's nondiscrimination policy was neutral towards religion and did not violate the agency's statutory or constitutional rights.[54] The Third Circuit affirmed, explaining that Philadelphia stood "on firm ground in requiring its contractors to abide by its non-discrimination policies," so long as the regulation was not "a veiled attempt to suppress disfavored religious beliefs."[55] Insofar as there was no evidence of anti-religious bias in the record, the Third Circuit held that the record reflected a good faith effort to enforce laws against discrimination.[56]

When the case was appealed to the Supreme Court, some legal commentators were quick to focus on "the attitude of the current

---

[49] *See, e.g.*, Joint Appendix Vol. I at 653–54, *Fulton*, 141 S. Ct. 1868 (No. 19-123), https://www.scotusblog.com/wp-content/uploads/2020/05/Fulton-v.-Philadelphia-19-123-Joint-Appendix-Volumes-I-and-II.pdf [https://perma.cc/8QTV-EREG].

[50] *See* Fulton v. City of Philadelphia, 320 F. Supp. 3d 661, 673 (E.D. Pa. 2018), *aff'd*, 922 F.3d 140 (3d Cir. 2019), *rev'd*, 141 S. Ct. 1868; Julia Terruso, *City Halts Foster Care Intakes at Two Agencies That Discriminate Against LGBTQ People*, Phila. Inquirer (Mar. 15, 2018), https://www.inquirer.com/philly/news/city-council-lgbtq-discrimination-foster-adopt-child-welfare-hearings-20180315.html [https://perma.cc/8S8S-XGW9].

[51] Complaint at 3–8, 24–39, *Fulton*, 320 F. Supp. 3d 661 (No. 18-cv-2075). The second agency with which Philadelphia cut ties did not join the suit.

[52] *See* Emp. Div., Dep't of Hum. Res. v. Smith, 494 U.S. 872, 879–82 (1990).

[53] *Id.* at 885–86 (stating that the "compelling government interest" requirement is not appropriate for generally applicable laws).

[54] *See Fulton*, 320 F. Supp. 3d at 684–85, 703.

[55] *Fulton*, 922 F.3d at 165.

[56] *Id.*

court" towards religion and sexual orientation discrimination.[57] All eyes were on Justice Barrett, who was confirmed just before oral argument.[58] In the end the Court reversed—unanimously—but on narrow grounds.[59] The majority held that the contract between the City and the agency provided the City with the authority to grant individualized exemptions from its nondiscrimination policy, and that refusing to do so meant that the City's actions were not neutral towards religion.[60] The City's policy was subjected to strict scrutiny and struck down,[61] all while leaving the big constitutional issues largely unaddressed.[62]

The major constitutional issues were instead brought to the stage in the three concurrences, two of which complained that the majority had failed to take on the conflict between free exercise and gay rights at the heart of the case. These concurrences suggested that the majority's decision to rule as narrowly as possible came at the price of misrepresenting both fact and law. Justice Gorsuch's concurrence described the majority opinion as engaging in a "dizzying series of maneuvers" which included an "uncharitably broad reading" of one provision of the contract and the rewriting of another.[63] Justice Alito explained that instead of confronting the constitutional question in the case, the majority's decision hung on a "glitch" in the contract.[64] Justices Alito and Gorsuch both opined that the Court's deciding the case so narrowly was a strategic move to avoid facing the question of whether *Smith* should be overturned.[65] Five Justices, including Justice Barrett, expressed willingness to overturn *Smith* and restore a more searching test for violations of the Free Exercise Clause.[66]

---

57 Erwin Chemerinsky, *SCOTUS Considers Whether Religious Freedom Also Means Freedom to Discriminate*, A.B.A. J. (Oct. 29, 2020, 10:27 AM), https://www.abajournal.com/columns/article/chemerinsky-scotus-considers-whether-religious-freedom-also-means-freedom-to-discriminate [https://perma.cc/3SGY-R9SL].

58 Michelle Boorstein, *Religious Conservatives Hopeful New Supreme Court Majority Will Redefine Religious Liberty Precedents*, WASH. POST (Nov. 3, 2020), https://www.washingtonpost.com/religion/2020/11/03/supreme-court-religious-liberty-fulton-catholic-philadelphia-amy-coney-barrettt [https://perma.cc/3CTF-KXJD] ("Barrett's addition to the court raises concerns among some that the balance will tilt further toward religious privilege for some at the expense of the nonreligious and religious minorities.").

59 *See* Fulton v. City of Philadelphia, 141 S. Ct. 1868, 1882 (2021).

60 *Id.* at 1878–79.

61 *Id.* at 1881–82.

62 *See id.*

63 *Id.* at 1929 (Gorsuch, J., concurring).

64 *Id.* at 1887–88 (Alito, J., concurring).

65 *See id.*; *id.* at 1929–30 (Gorsuch, J., concurring).

66 *See id.* at 1883, 1926 (Alito, J., concurring) (explaining that *Smith* is "ripe for reexamination" and that he "would overrule *Smith*"); *id.* at 1931 (Gorsuch, J., concurring) (noting that no Justice defended *Smith* in response to Justice Alito's concurrence and asking

Fulton resembles *Masterpiece Cakeshop v. Colorado Civil Rights Commission*, an LGBTQ equality versus religious freedom case decided just a few years earlier.[67] *Masterpiece Cakeshop* involved a baker who refused to create a wedding cake for a same-sex couple, citing religious beliefs.[68] The majority sidestepped the constitutional questions at the heart of the conflict and decided the case on procedural grounds.[69] In his concurrence, however, Justice Thomas advanced his view that the baker should have prevailed on constitutional grounds.[70] And Justice Ginsburg's dissent contended that the procedural issues should not have "overcome [the baker's] refusal to sell a wedding cake" to the couple.[71] Again, some viewed the decision as narrow,[72] while others worried about its breadth.[73] The acute disagreement about the likely consequences of these decisions illustrates the importance of measuring them.

The most recent clash between religion and equality, *303 Creative LLC v. Elenis*,[74] again involved a wedding vendor (this time, a website designer) seeking an exemption from antidiscrimination obligations. In a narrowing maneuver, the Court set itself up to avoid the religious freedom issue entirely, by granting certiorari only on the free speech question.[75] Nonetheless, the in-Court fighting revealed very different

what the Court is "waiting for"); *id.* at 1882 (Barrett, J., concurring) (noting the "textual and structural arguments against *Smith*").

67 Masterpiece Cakeshop, Ltd. v. Colo. C.R. Comm'n, 138 S. Ct. 1719 (2018).

68 *Id.* at 1723.

69 *See id.* at 1732 (holding that the adjudicating body was hostile towards religion rather than directly deciding whether religious liberty should prevail over LGBTQ equality).

70 *Id.* at 1740–42 (Thomas, J., concurring) (arguing that the baker's conduct was expressive and should be protected under the First Amendment).

71 *Id.* at 1751 (Ginsburg, J., dissenting).

72 *See, e.g.*, Erwin Chemerinsky, *Not a Masterpiece: The Supreme Court's Decision in* Masterpiece Cakeshop v. Colorado Civil Rights Commission, 43 Hum. Rts., no. 4, 2018, at 11 (noting that the Court did not decide whether religious freedom was more important than protecting equality, but "instead decided the case on narrower grounds").

73 *See, e.g.*, Emilie Kao, *Why the Supreme Court's Ruling for a Christian Baker Was Not "Narrow,"* Heritage Found. (June 12, 2018), https://www.heritage.org/courts/commentary/why-the-supreme-courts-ruling-christian-baker-was-not-narrow [https://perma.cc/5DSD-J4ZZ] ("The court's clear rejection of the discrimination argument has implications for many of the other conflicts currently brewing between religious freedom and sexual orientation."); Mark Hemingway, *Why* Masterpiece Cakeshop *Is a Win for Religious Freedom*, Wash. Exam'r (June 4, 2018), https://www.washingtonexaminer.com/weekly-standard/supreme-court-why-masterpiece-cakeshop-is-a-win-not-just-for-jack-phillips-but-also-religious-freedom [https://perma.cc/5CP4-XY67] (noting that the decision "represents a significant victory for proponents of religious religious [sic] liberty and freedom of conscience").

74 303 Creative LLC v. Elenis, 143 S. Ct. 2298 (2023).

75 *Compare* 303 Creative LLC v. Elenis, 6 F.4th 1160, 1168 (10th Cir. 2021) (deciding the litigation on both free speech *and* free exercise grounds, among other issues), *rev'd*, 143 S. Ct.

interpretations of the breadth of the decision—that again favored the religious party. Justice Gorsuch, writing for the Court, insisted that the majority's decision was consistent with precedent and did not expand discrimination.[76] Justice Sotomayor, on the other hand, wrote in her dissent that the decision provided a broad license to discriminate and marked "a sad day in American constitutional law and in the lives of LGBT people."[77] Readers might conclude, wrote Justice Gorsuch, that "we are [not] looking at the same case."[78]

These cases came to the Court after a generation-long ferment in the antidiscrimination space. Specifically in foster care (though also in public accommodations law) an increasing number of states—totaling thirty-five by 2021—enacted laws protecting prospective parents against discrimination on the basis of sexual orientation.[79]

Many of these states enacted these rules specifically for foster care agencies, while other states applied their general public accommodation laws or antidiscrimination rules to this context.[80] In most cases, sexual orientation protections were introduced to existing antidiscrimination laws that already protected from other forms of discrimination, expanding their scope.[81] The regulation of the foster care system is largely

---

2298, *with* 303 Creative LLC v. Elenis, 142 S. Ct. 1106 (2022) (granting certiorari only on the free speech question in the case).

76 *303 Creative*, 143 S. Ct. at 2314–18.

77 *Id.* at 2341 (Sotomayor, J., dissenting).

78 *Id.* at 2318 (majority opinion).

79 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, 1015 (2022). We rely on the mapping conducted by Netta Barak-Corren, Yoav Kan-Tor, and Nelson Tebbe, who collected data on the date each state introduced an antidiscrimination or religious exemption rule providing or excluding protections for prospective same-sex couples in foster care. *See id.* at 1003. To that end, Professor Barak-Corren and her colleagues analyzed all legal provisions, whether grounded in statute, regulation, state agency policy, or court decision, and also consulted the databases of antidiscrimination laws compiled by the nonprofit organizations Movement Advancement Project and Lambda Legal (as these databases looked at the time of the study). They resolved ambiguities through consultation with family law experts. *Child Welfare Nondiscrimination Laws*, Movement Advancement Project, https://www.lgbtmap.org/equality-maps/foster_and_adoption_laws [https://perma.cc/Q93A-HDN4]; *Child Welfare*, Lambda Legal, https://www.lambdalegal.org/map/child-welfare [https://web.archive.org/web/20220701051942/https://www.lambdalegal.org/map/child-welfare].

80 For example, New Mexico's nondiscrimination statute generally prohibits discrimination on the basis of sexual orientation and gender identity, *see* N.M. Stat. Ann. § 28-1-7 (2023), and the state also has promulgated more specific regulatory protections against discrimination on these bases by foster agencies, *see* N.M. Code R. § 8.26.5.15 (LexisNexis 2023).

81 For an example of a general nondiscrimination law that was expanded to include sexual orientation, see 110 Mass. Code Regs. 1.09(3) (2000), https://archives.lib.state.ma.us/bitstreams/45ccf905-ad2a-4ad3-901d-d48836d7f6af/download [https://perma.cc/P3PB-QR8K] (forbidding sexual orientation discrimination against recipients of any

centralized at the state level, meaning antidiscrimination rules in this setting are typically state-wide. Philadelphia's *local* antidiscrimination rules—the subject of the dispute in *Fulton*—are the exception rather than the norm in this respect.[82]

At the same time, other states—eleven by 2021—enacted laws providing religious foster care agencies a right to *deny* service to same-sex couples, commonly referred to as religious exemption laws.[83] Three states provided this right against an otherwise applicable prohibition on sexual orientation discrimination,[84] and the remaining eight had never prohibited foster care agencies from discriminating based on sexual orientation, but nevertheless preemptively enacted religious refusal rights.[85]

Figure 1 illustrates the state of foster care antidiscrimination law in all fifty states and the District of Columbia before *Fulton*. The Figure shows a "legislative mismatch"[86] between states with antidiscrimination

services of the Massachusetts Department of Social Services). For a licensing rule with a nondiscrimination requirement, see 102 Mass. Code Regs. 1.03(1) (1997), https://archives.lib.state.ma.us/bitstreams/ad46e31f-6aa1-4365-b7ee-801fd7578994/download [https://perma.cc/LDE2-32MA] (requiring private foster care agencies, as licensees of the department, to not "discriminate in providing services to children and their families on the basis of . . . sexual orientation"). We focus on prohibitions of discrimination against prospective parents, as opposed to youth.

82 Barak-Corren, Kan-Tor & Tebbe, *supra* note 79, at 1015 (finding that most states administer the child welfare system at the state level, nine states including Pennsylvania administer the child welfare system wholly at the county level, and two states partially administer it at the county level). In other settings—public accommodation laws, for example—it is much more common for local governments to enact their own non-discrimination rules. *See* Netta Barak-Corren, *A License to Discriminate? The Market Response to* Masterpiece Cakeshop, 56 Harv. C.R.-C.L. L. Rev. 315, 322, 328–30 (2021) [hereinafter Barak-Corren, *A License to Discriminate?*] (noting that "a considerable number of local governments . . . enacted municipal [antidiscrimination] laws").

83 Barak-Corren, Kan-Tor, & Tebbe, *supra* note 79, at 1015. *Fulton* changed this landscape; for example, Arizona's administrative code explicitly provides that foster care agencies cannot discriminate against prospective parents on the basis of sexual orientation. *See* Ariz. Admin. Code § R21-6-201 (2023). In 2022, after *Fulton*, Arizona carved out a religious objection exemption by statute that expressly overrode that regulation. *See* S.B. 1399, 55th Leg., 2nd Reg. Sess. (Ariz. 2022).

84 *See, e.g.*, Mich. Comp. Laws Ann. § 710.23g (2015).

85 For an example of a general religious refusal law, see Miss. Code Ann. §§ 11-62-3, 11-62-5 (2023). For an example of a foster care-specific religious refusal law, see N.D. Cent. Code §§ 50-12-03, 50-12-07.1 (2023). For an example of a state religious freedom restoration act, see R.I. Gen. Laws § 42-80.1-3 (2023). Notably, all of the religious refusal laws in the foster care context are specific. Unlike the more general Religious Freedom Restoration Laws, which only provide religious individuals and organizations the possibility of securing an exemption once all the conditions of the law are met, religious refusal laws are phrased as guarantees of the right to refuse service, and prohibit governments from any adverse action against an agency for any such refusal.

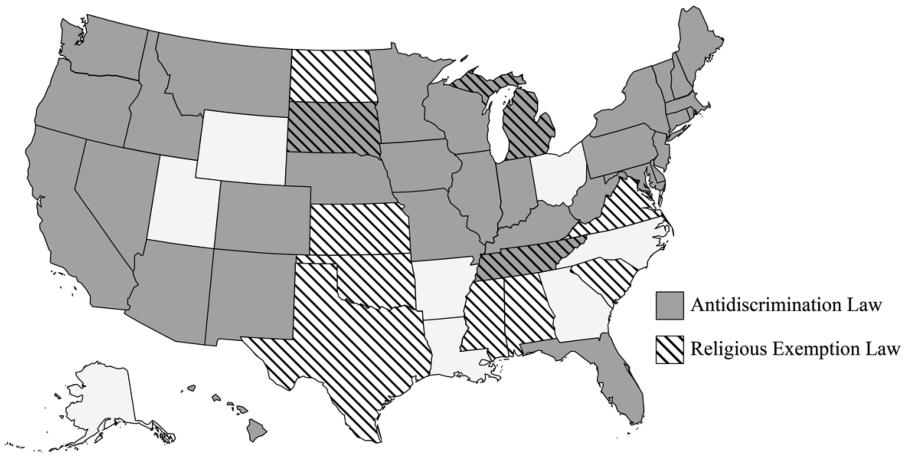
86 This term was first used by Ira Lupu in the context of the mismatch between public accommodation laws protecting against sexual orientation discrimination and Religious

laws and states with religious exemption laws, with very narrow overlap. There are four basic types of states:

TABLE 1. STATE ANTIDISCRIMINATION AND RELIGIOUS EXEMPTION TYPOLOGIES

	No Antidiscrimination Rules	**Antidiscrimination Rules**
No Religious Exemption Rules	*Neither*: Eight states that enacted no antidiscrimination laws and no religious exemption laws (e.g., Ohio).	*Antidiscrimination Only*: Thirty-two states that enacted sexual orientation antidiscrimination rules (and no religious exemption rules) (e.g., Minnesota).
Religious Exemption Rules	*Religious Exemption Only*: Eight states that enacted only religious refusal rules (and no antidiscrimination rules) (e.g., Texas).	*Both*: Three states that enacted both antidiscrimination rules and religious exemption rules (e.g., Michigan).

FIGURE 1. STATE ANTIDISCRIMINATION AND RELIGIOUS EXEMPTION LAWS FOR FOSTER CARE AGENCIES, 2021



Freedom Restoration Laws. *See* Ira C. Lupu, *Moving Targets:* Obergefell, Hobby Lobby, *and the Future of LGBT Rights*, 7 ALA. C.R. & C.L. L. REV. 1, 48–49 (2015). Notably, the legislative mismatch of the foster care antidiscrimination laws does not track the mismatch of the public accommodations antidiscrimination laws.

To make this map concrete, consider a foster care agency that has religious objections to working with a same-sex couple. In states like Minnesota, the law prohibits the agency from discriminating on the basis of sexual orientation, and no religious exemption is available. There, religious agencies will have to (at least) declare they abide by the law to keep their licenses. In contrast, in states like Texas, objecting religious agencies are free to refuse service to same-sex couples. The government may not revoke their licenses and legal claims against them are barred. This is likely the case in states like Ohio, which enacted neither type of law. Absent relevant antidiscrimination rules, agencies are likely free to refuse service under the law. In the overlap category, where both sexual orientation and religious refusal are afforded legislative protections, objecting agencies are similarly granted the right to refuse service and are free to do so.

This patchwork is the background against which *Fulton* was decided, and against which the debate on religious exemptions is still raging.

II
The Field Experiment

When the Supreme Court granted certiorari in *Fulton v. City of Philadelphia* and signaled willingness to intervene in the nondiscrimination space and potentially expand religious exemptions, we saw an opportunity to conduct a field experiment examining the consequences of this constitutional move—and we seized it.

As *Fulton* dealt with one of the most salient and recurring battles in constitutional law—that between religion and equality—and could have far-reaching implications for the entire foster care field,[87] it was greatly anticipated in these circles. It attracted eighty-eight amicus briefs[88] and drew wide interest from foster care groups,[89] religious

[87] *See* Madeleine Carlisle & Belinda Luscombe, *The Most Powerful Court in the U.S. Is About to Decide the Fate of the Most Vulnerable Children*, Time (May 28, 2021, 2:02 PM), https://time.com/6051046/supreme-court-foster-care-fulton-philadelphia [https://perma.cc/YLJ6-ZUNT] (describing how the Court's decision could affect the rights of LGBTQ families to adopt and the number of operating foster organizations).

[88] *Fulton v. City of Philadelphia, Pennsylvania*, SCOTUSblog, https://www.scotusblog.com/case-files/cases/fulton-v-city-of-philadelphia-pennsylvania [https://perma.cc/82XK-ARRL].

[89] *See, e.g.*, Brief for FosterClub & Former Foster Youth as Amici Curiae Supporting Respondents, Fulton v. City of Philadelphia, 141 S. Ct. 1868 (2021) (No. 19-123); Candice Dundy, *Fulton v. City of Philadelphia*, Legal Counsel for Youth & Children (Oct. 15, 2020), https://lcycwa.org/blog/2020/10/13/fulton-v-city-of-philadelphia [https://perma.cc/LX3K-JWM9] (detailing possible effects of *Fulton* and how the Legal Counsel for Youth and Children signed on to an amicus brief in support of Philadelphia).

organizations and outlets,[90] and LGBTQ groups.[91] Entire conferences were dedicated to discuss its potential implications.[92] That the case received so much attention, especially from stakeholders who would immediately conform their actions to the holding of the case, made it an especially good candidate for a field experiment.

A. An Introduction to Field Experiments in Law, and Our Design

Field experiments provide the most powerful method for causal inference in real life settings currently known to the social and legal sciences.[93] In particular, situating the measurement of the treatment accorded to same-sex and opposite-sex couples directly around the *Fulton* decision allows us to study the implications of an actual, concrete, and important legal event in a controlled setting.

---

[90] *See, e.g.*, Brief for Galen Black as Amici Curiae Supporting Petitioners, Fulton v. City of Philadelphia, 141 S. Ct. 1868 (2021) (No. 19-123) (showing an amicus brief written by the First Liberty Institute in support of the petitioners); Fulton v. City of Philadelphia, FREE TO FOSTER, https://freetofoster.com/case-details [https://perma.cc/2TJN-QUAH] (showing coverage of *Fulton* from a foster care blog supporting the petitioners in the case); Andrea Picciotti-Bayer, *Supreme Court* Fulton *Case Will Soon Decide Fate of Faith-Based Foster Care*, NAT'L CATH. REG. (Oct. 29, 2020), https://www.ncregister.com/commentaries/supreme-court-fulton-case-will-soon-decide-fate-of-faith-based-foster-care [https://perma.cc/T89Y-87DC] (providing Catholic-perspective commentary on *Fulton* in favor of petitioners); Carol Zimmerman, *All Eyes on U.S. Supreme Court as Religious Liberty Case from Philadelphia Foster Care Program Set for November Argument*, DIALOG (Sept. 2, 2020), https://thedialog.org/national-news/all-eyes-on-u-s-supreme-court-as-religious-liberty-case-from-philadelphia-foster-care-program-set-for-november-argument [https://perma.cc/7JLZ-2S8D] (providing Catholic-perspective news coverage of *Fulton*).

[91] *See, e.g.*, Anagha Srikanth, *The Supreme Court's Next Major Case Starts the Day After the Election*, HILL (Oct. 6, 2020), https://thehill.com/changing-america/respect/equality/522865-the-supreme-courts-next-major-case-starts-the-day-after-the [https://perma.cc/8AJH-N34D]; Fulton v. City of Philadelphia*: Amicus Brief*, WILLIAMS INST., https://williamsinstitute.law.ucla.edu/publications/fulton-v-city-of-philadelphia-amicus-brief [https://perma.cc/DW4D-WMQC] (showing coverage of *Fulton* from an institute focused on sexual orientation and gender identity issues).

[92] *See, e.g.*, *Symposium Before Oral Argument in* Fulton v. City of Philadelphia, SCOTUSBLOG, https://www.scotusblog.com/category/special-features/symposia-before-oral-arguments-in-the-2020-21-term/symposium-before-oral-argument-in-fulton-v-city-of-philadelphia [https://perma.cc/Q2YU-GFJ3] (collecting various symposia on *Fulton*); Fulton*, Faith, Families, & Foster Care Symposium*, GEO. WASH. UNIV., https://www.law.gwu.edu/fulton-faith-families-foster-care-symposium [https://perma.cc/8HFH-C2TB] (showing the schedule for a symposium on *Fulton* held virtually on January 22, 2021 at the University of Virginia). Family Court Review dedicated a special issue to articles presented in this symposium. *See* Symposium, *Special Issue: Foster Care, Same Sex Couples, and the Constitution*, 60 FAM. CT. REV. 1 (2022).

[93] *See, e.g.*, Marianne Bertrand & Sendhil Mullainathan, *Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination*, 94 AM. ECON. REV. 991, 993–94 (2004) (discussing the advantages of field experiments over other forms of experiments).

Field experiments have their limitations, too. Ensuring a controlled experiment in field settings requires intricate and detailed coordination and typically requires researchers to simplify their design to preserve its integrity. Therefore, we are typically limited in the number of interventions we are able to test in any given experiment.[94] In addition, the measurement of real behavior involves particular ethical concerns. Researchers need to take into account the extra work they create for the decisionmakers whose decisions are measured. They also must consider the lack of ability to obtain informed consent, due to the need to observe the behavior in its natural settings. As such, ethical considerations may require researchers to engage in more light-touch designs that are more limited in the range of behaviors they can measure (as we did here).[95] Notably, while other types of empirical investigation can teach us important lessons—for example, observing correlations between variables can teach us about patterns and relationships in the real world, and lab experiments can isolate causal mechanisms in controlled and sterilized settings—only field experiments combine the advantages of observing real behavior in a natural setting with the advantages of strong causal inference afforded by randomization and controls. If one wants to know whether foster care agencies discriminate between same-sex and opposite-sex couples, and whether such behavior is influenced by a major legal event in the field such as *Fulton*, there is no better way to find out than to conduct a field experiment.

We designed our field experiment to assess whether foster care agencies discriminate against same-sex couples, whether discrimination varies by jurisdiction type and agency type, and to what extent the *Fulton* decision influenced discrimination patterns. We measured the behavior of foster care agencies in two periods: before and after June 17, 2021, the day that *Fulton* was handed down. We sent emails to every foster care agency in the country, randomly assigning them to be from

[94] For this reason, which influences statistical power analysis, we only examine willingness to provide service to male couples and do not examine impacts on lesbian or non-binary couples or couples with distinctively Black or non-white names, nor do we explore the intersectionality of gender and race. These are all important topics for future studies. However, if we attempted to examine all or even some of these questions in this study, we would be under-powered to detect an effect. *See* Mattie Mackenzie-Liu, David J. Schwegman & Leonard M. Lopoo, *Do Foster Care Agencies Discriminate Against Gay Couples? Evidence from a Correspondence Study*, 40 J. Pol'y Analysis & Mgmt. 1140, 1149 n.11 (2021) (describing lack of statistical power to detect statistically significant results in a field study with inquiries from same-sex male, same-sex female, and opposite-sex couples).

[95] The study was reviewed and approved by the Hebrew University IRB (March 2021). We have also pre-registered our design in aspredicted.org under pre-registration No. 64948. *Fulton Field Experiment (#64948)*, Wharton Credibility Lab, https://aspredicted.org/3ug6p.pdf [https://perma.cc/Y7F5-NX6R].

a same-sex or an opposite-sex couple, and then coded the valence of any reply to that email. We matched each agency with its jurisdiction type, tracing whether it is subject to antidiscrimination protections for prospective gay parents and/or whether it is entitled to refuse service to prospective gay parents.

This design has some important advantages for causal inference and external validity. First, ours is a classic audit study, made stronger because we have a before and after measurement that allows us to track the causal impact of the Court's decision.[96] Second, unlike most studies that draw on a sample of the population of interest, this study covers the entire population of foster care agencies in the United States. Relatedly, we have collected data from all fifty states. These strengths allow us both to generalize from the results and provide state-level estimates of discrimination, which have never before been measured. Unusually, we were also able to directly identify foster care agencies' religiosity, an element missing from the vast majority of previous studies of sexual orientation discrimination and a crucial piece of the analysis of the dynamics of religion-equality conflicts.[97] Our study is the first to provide

[96] For an example of a classic audit study, see Bertrand & Mullainathan, *supra* note 93. The authors sent otherwise identical resumes with racially coded names in response to help-wanted ads and found that African-American candidates got "far fewer callbacks for each resume they sen[t] out." *Id.* at 1011. Amanda Agan and Sonja Starr's "Ban the Box" study was the first study in law to combine an audit with a before-after measurement. *See generally* Amanda Agan & Sonja Starr, *Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment*, 133 Q.J. Econ. 191 (2018). The authors used an audit study to measure the impact of "Ban the Box" laws that prohibit potential employers from asking about applicants' criminal history, sending fake online job applications on behalf of young men to entry-level roles before and after the laws went into effect. *Id.* at 192–93. For a subsequent adaptation of these methods to the religion/equality conflict, see generally Barak-Corren, *A License to Discriminate?*, *supra* note 82 (utilizing an audit study combined with a before-after measurement to assess the impact of *Masterpiece Cakeshop* on LGBTQ discrimination in the wedding vendor market).

[97] In a previous field experiment examining discrimination against same-sex couples in the wedding market, Professor Barak-Corren did not have direct identification of religiosity and measured instead the religiosity of the business environment. *See* Netta Barak-Corren, *Religious Exemptions Increase Discrimination toward Same-Sex Couples: Evidence from* Masterpiece Cakeshop, 50 J. Legal Stud. 75, 101–02 (2021). Neither was religiosity identified in previous studies examining the effects of Supreme Court decisions on public attitudes towards same-sex marriage. *See* Margaret E. Tankard & Elizabeth Levy Paluck, *The Effect of a Supreme Court Decision Regarding Gay Marriage on Social Norms and Personal Attitudes*, 28 Psych. Sci. 1334 (2017). The only exception, as far as we know, is a recent paper by Mattie Mackenzie-Liu, David J. Schwegman, and Leonard M. Lopoo that reanalyzes data from a 2018 correspondence study of sexual orientation discrimination by foster care agencies, this time coding agencies' religiosity. *See* Mattie Mackenzie-Liu, David J. Schwegman & Leonard M. Lopoo, *Do Faith-Based Foster Care Agencies Respond Equally to All Clients?*, 37 J. Pol'y Stud. 44 (2022) (reanalyzing data from their 2018 study, *Do Foster Care Agencies Discriminate Against Gay Couples? Evidence from a Correspondence Study*, by coding for agencies' religiosity).

comprehensive and recent evidence of agencies' behavior before and after the *Fulton* decision, by states, legal regime, and the share of religious agencies in the state.

1. The Sample

We assembled the sample of foster care agencies ("agencies") in two stages. First, we scraped the websites of all states and built a database that was intended to cover all agencies providing services to prospective foster parents across the fifty states (for convenience, we will refer to this database as the "sample," though it aimed to cover the entire population of interest).[98] The dataset contained the name of the agency, its location, and contact information. At this stage, agencies with names indicative of religious affiliation (e.g., Catholic Social Services; Bethany Christian Services; Baptist Child & Family Services) were automatically identified as religious agencies.

Next, we manually verified the dataset by visiting each agency's website. This served three purposes. First, we verified the agency's contact information—preferably an email address, and in the absence of one, an online contact form with a free text inquiry box. Second, we verified the agency's type as public (state or county department of social/human services), religious, or non-religious. In this process, agencies whose websites had mentions of religion, faith, or other indicators of religiosity that were not evident from their names were added to the list of religious agencies. Third, we refined the dataset by removing duplicate entries and agencies that turned out to be inactive or irrelevant for the purposes of our study (mostly agencies that only provided adoption services);[99] by adding missing branches to networks for which only some of the branches originally appeared in the dataset; and by unifying into one entry different branches of a network that shared identical contact information.

[98] Every state has a website, typically maintained by its department of human or home services, that lists all of the foster care agencies licensed to operate in that state. With the help of the Hebrew University Center for Interdisciplinary Data Science Research (CIDR), we scraped these websites and created a database of agencies.

[99] This process was not exhaustive: after fielding the experiment, we discovered additional agencies that were adoption-only or inactive and excluded them from the sample. We also eliminated agencies for which Gmail could not deliver our inquiry to their email addresses.

Our final sample comprises 1,235 agencies in the experimental group and 220 agencies in the control group: 1,455 agencies (public,[100] private non-religious, and religious) with complete observations.[101]

2. *Procedure*

We created eight fictitious e-mail profiles for the experiment, representing persons who are part of a same-sex or an opposite-sex couple.[102] To collect baseline levels of discrimination, we sent each agency two inquiries—one from each couple type—before the Court's decision in *Fulton* (over the course of March 2021: waves 1 and 2) and two such inquiries after the decision (in the first and second week after the decision: waves 3 and 4).[103] The order of the two inquiries was randomized in the first wave in each period, such that half of the agencies received emails from same-sex couples and half from opposite-sex couples (and the other way around in the following wave). Altogether, every agency received four emails from four different characters, two from a same-sex couple and two from an opposite-sex couple. We block randomized the allocation of sexual identity within each group of agencies (public, religious, and non-religious) and within states. The 220 agencies of the control group were randomly allocated to be contacted only after the *Fulton* decision.

We also used a longitudinal survey (that we describe below) to monitor the stability of public opinion regarding religious agency refusal to same-sex couples during the waiting period from March to June 2021. Each month, we issued a survey to about 300 participants and asked them about their support for faith-based foster care agencies refusing service to same-sex couples. We found no significant changes in public opinion on this topic during late March/early April (dates corresponding

---

[100] The sample includes public agencies from all states except five, for which we were not able to find online contact information. This happened for two reasons. First, some states' websites could not be entered into by individuals from out-of-state locations (Florida and Maryland). Second, some states' websites did not publish contact information other than phone numbers (Texas, Iowa, and Maine). Because those states require prospective parents to call for more information, they did not fit into the parameters of our email-driven study. In Hawaii, Oregon, and Wyoming, we found only public agencies, and no private agencies operating foster parent services.

[101] Our initial dataset contained 1,905 foster care agencies. The Appendix discusses the reasons for missing observations and exclusions, including blocked and bounced emails and cases of suspicion. *See infra* app. I–J.

[102] Eight additional profiles were created following the same procedures for Bethany Christian Services foster care network, after we learned in wave 1 that the network uses a shared database of inquirers, thereby allowing agents to see if a person contacted branches in different geographical locations. *See infra* app. H tbl.A7.

[103] The dates were as follows: 1st wave—March 15–22 (8 days); 2nd wave—March 29–31 (3 days); 3rd wave—June 23–26 (4 days); 4th wave—June 29–July 3 (5 days).

to those of our pre-*Fulton* email inquiries), early May, and early June. Public opinion remained stable from our pre-*Fulton* measurement until right before *Fulton*, and changed (as we report below) only after *Fulton* was handed down.[104] This monitoring provides further assurance that the results we observe in both the field experiment and the longitudinal survey are not due to some other event that preceded *Fulton*.[105]

The sexual orientation of the couple was signaled through the name and description of the sender's spouse (wife/husband).[106] The senders had common male names (John, Robert, David, Scott) which were apparent in their Gmail profile and email signatures. The signatures also included the sender's phone number to increase the profiles' reliability.[107]

We drafted the emails based on foster care agencies' online materials, including information for prospective parents, help pages, and FAQ pages.[108] We kept the content of the emails identical across couples' sexual orientations, such that the only distinction between the emails was the gendered names of the spouses of the applicant. Each agency received a different email, from a different couple, in each wave. Each of these emails contained an expression of interest in becoming foster parents and short requests that required only a brief and simple reply from the agency's representative (e.g., providing the date of an upcoming orientation, a number to call to talk to a representative, or a contact to schedule a meeting). The emails also included a brief introduction of the couple, including some personal information, including age (around forty), profession (stable and not time-consuming jobs), a spouse profession (stable full-time jobs), children (either one

---

104 *See infra* Section II.B.

105 There was a dense cluster of religious exercise Supreme Court decisions immediately preceding *Fulton*, brought on by litigation surrounding COVID-19 restrictions. *Roman Catholic Diocese of Brooklyn v. Cuomo*, 141 S. Ct. 63, 69 (2020), decided in November 2020, enjoined pandemic restrictions for religious service gatherings. In February 2021, the Court partially lifted California's restrictions on religious services. *See* S. Bay United Pentecostal Church v. Newsom, 141 S. Ct. 716 (2021). *Tandon v. Newsom*, 141 S. Ct. 1294, 1297 (2021), granted an injunction blocking COVID-19 restrictions from interfering with at-home religious gatherings in April 2021. Of these cases, only *Tandon* could have interfered with our measurements, but our monthly surveys empirically ruled out this possibility. *See infra* Section II.B.3.

106 Female spouses were named Nicole, Sarah, Lisa, or Amanda and male spouses were named James, Charles, Paul, or Harry.

107 The numbers were not available for response, but we were able to see missed calls and text messages.

108 For examples of FAQ pages, see *Foster Care Frequently Asked Questions*, Wis. Dep't of Child. & Fams., https://dcf.wisconsin.gov/fostercare/faq [https://perma.cc/VS9N-GF45]; *Foster Parent FAQ*, NECCO, https://www.necco.org/foster-parent-faq [https://perma.cc/P5NB-SBLX]; *Becoming a Foster Parent: FAQs*, HealthyChildren.Org, https://www.healthychildren.org/English/family-life/family-dynamics/adoption-and-foster-care/Pages/Foster-Parents-FAQs.aspx [https://perma.cc/3WSH-49BA].

or two), and for some, a pet.[109] Our preparatory research found these characteristics reflect an attractive foster family candidate.[110] Here is how one of our emails read[111]:

> Hi,
>
> My name is [Name], and my [husband/wife] [husband/wife Name] and I are considering to become foster parents. We understand that your agency can help us go through this process and would like to hear some more. I work as a part time soccer team coach and [husband/wife Name] is a human resource manager. We are currently approaching our 40's and having met in high school, we have been together for quite a while now. Together, we have built a nice little family with two beautiful children.
>
> Our main question regarding foster care and taking another child into our home is the response of our children. We saw some articles discussing this matter online, but we are not sure about their accuracy. We were wondering if we can talk to someone or whether there is an orientation that addresses these and other questions we might have?
>
> Best wishes,
> [Name]

To respect the replies of the agencies and reduce suspicion and fatigue throughout the experiment, the research team answered each responding agency individually with a polite and friendly thank you email soon after the response was received and before the next wave of inquiries was sent.

3. Coding the Data

We received 2,135 response emails which were coded according to a detailed protocol and observed 1,570 instances of no response,

---

[109] The third wave email versions alternated between including one of two aspects of personal information: age or occupation. In the fourth wave, no biographical information was included, to avoid similarly worded requests that may raise suspicion over time. Unfortunately, as we discuss in the appendix, suspicion arose nevertheless in that wave. *See infra* app. J.

[110] *See, e.g.*, SHARON VANDIVERE, KARIN MALM & LAURA RADEL, U.S. DEP'T OF HEALTH & HUM. SERVS., ADOPTION USA: A CHARTBOOK BASED ON THE 2007 NATIONAL SURVEY OF ADOPTIVE PARENTS 16 (2009); Andrew Zinn, *Foster Family Characteristics, Kinship, and Permanence*, 83 SOC. SERV. REV. 185, 208–11 (2009) (discussing statistical research findings regarding rates of adoption amongst foster placements). While some of these findings relate to characteristics that make for attractive adoptive family candidates, we assumed that these qualities—indicative of family stability and functionality—would likewise be desirable in foster family candidates.

[111] All of our emails are included in the Appendix. *See infra* app. D.

a total of 3,705 observations.[112] Six research assistants coded the entire data of email messages under the authors' close supervision. The research team conducted weekly meetings throughout the coding process to discuss the coding method, resolve open issues, and refine the coding scales.[113]

The outcome of interest in all our analyses is whether foster care agencies agreed to provide services to the couple. We applied two measures to this outcome, nuanced and binary. The scale for a nuanced response included the following: a positive response (i.e., an explicit willingness to provide service); a non-rejection (i.e., uninviting answers or answers that only implicitly conveyed willingness to provide service); no response; and a negative response (i.e, a refusal). The binary response unified the four nuanced response levels into two: no-response or negative response were coded as 0 and positive response and a non-rejection were coded 1. For ease of interpretability, we focus on the binary response measure in all our analyses, but the nuanced response measure yielded the same results.[114]

B. Results

1. Baseline Discrimination

When we contacted them in March 2021—pre-*Fulton*—foster care agencies were more likely to respond to emails from opposite-sex couples than same-sex ones. Overall, 60.9% of the agencies responded positively to (say) David and Amanda, but only 54.3% to David and Paul, a 12.2% difference.[115]

States that had previously enacted antidiscrimination laws without religious exemptions showed insignificant differences between couples (a 3% gap over 743 agencies and 1486 contacts). By contrast, states that enacted religious exemptions to their antidiscrimination laws showed large and significant differences (18% gap over 67 agencies

---

112 For a discussion of data exclusions, see *infra* app. I.

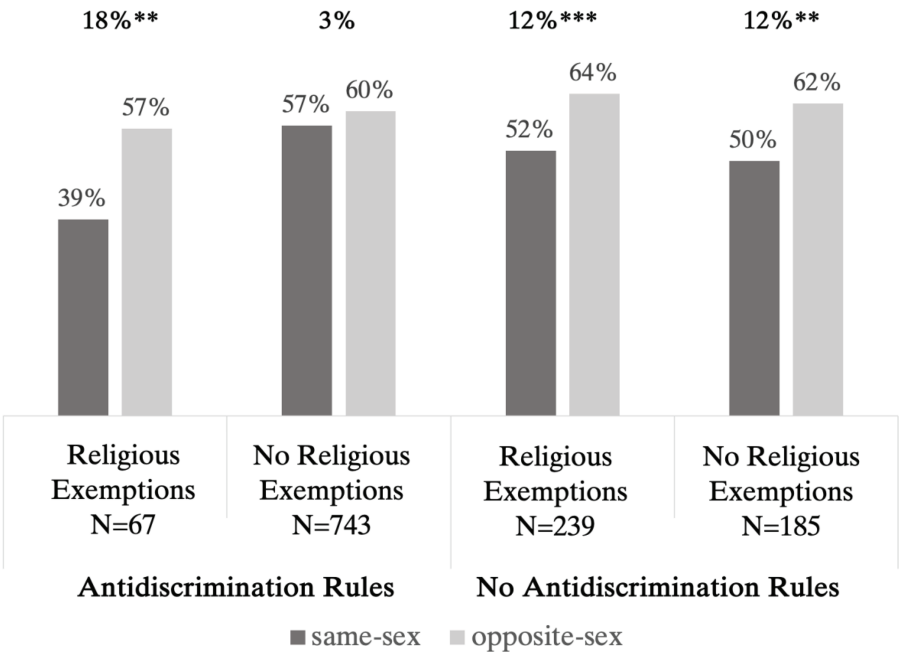
113 We employed an agency-oriented coding process. Each coder was allotted a random subset of agencies to code and coded the responses from all four waves of a specific agency at a time, giving the coder a wide perspective on the communication with the agency and a comprehensive understanding of how the agency functions. Agencies that were part of a network (including, for that matter, public agencies of the same state) were allocated to the same coder. We viewed the public agencies of a specific state as a network as well. The Appendix provides further information about the coding process. *See infra* app. F.

114 The number of incidences where Nuanced Response equals -1 or 0.5 is small. Therefore, analyses using the Nuanced Response variable reached the same results.

115 The difference is statistically significant ($p < .001$, N = 1,235 agencies multiplied by two waves, yielding 2,470 observations; agency fixed effects included in all analyses).

and 134 contacts). States that did not enact antidiscrimination laws showed a 12% gap, regardless of whether they also enacted religious exemptions.[116] Figure 2, below, illustrates the findings.

FIGURE 2. POSITIVE RESPONSE RATES OF FOSTER CARE AGENCIES TO SAME-SEX AND OPPOSITE-SEX COUPLES BY LEGAL REGIME BEFORE FULTON



* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

We also find that different types of agencies vary considerably in their willingness to provide service to couples, across sexual orientations. Public agencies provided the highest rates of positive responses, 64% on average, whereas private non-religious agencies provided 4.9% fewer positive responses on average and religious agencies provided 21%

[116] With no religious exemptions: $p < 0.05$; N = 185 agencies x two waves, yielding 370 observations; with religious exemptions: $p < 0.01$; N = 239 agencies x two waves, yielding 478 observations. It is also noticeable that the four regimes are not equal in size: most agencies (743) are located in antidiscrimination regimes that provide no religious exemptions, the group for whom *Fulton* is most relevant. Sizeable groups of agencies are located in regimes without antidiscrimination laws (424 in total) and far fewer agencies are located in regimes that enacted both antidiscrimination laws and exemptions from these laws for religious agencies.

fewer positive responses than public agencies and ~16% fewer positive responses than non-religious private agencies.[117] This finding suggests that religious agencies are generally less willing to provide service to couples than public and private non-religious agencies.

Overall, none of this should be terribly surprising, and it is in line with some of the other literature.[118] Simply put, jurisdictions that have enacted laws that are protective of gay rights (and concomitantly averse to religiously linked discriminatory practices) are also ones where foster agencies appear to be more up for treating gay and straight couples alike, at least at the door. We want to offer some cautions about this result. First, the fact that the behavior of foster care agencies varies greatly between legal jurisdictions does not mean that state law necessarily *caused* the observed differences. These laws—unlike the experimental treatment of sexual orientation—are not randomly assigned. In addition, unlike the *Fulton* decision, they are not new, so their effect cannot be studied as a natural before-after experiment (sometimes researchers have access to longitudinal data that allows for such study, but not in our case).

While no law can ever be easily separated from the underlying political and social climate that produced it, our data did allow us to control for various socio-political variables in all of the jurisdictions we surveyed. For each state, we collected data on its GDP and survey information on the average importance of religion, its share of conservatives, its share of Evangelical Christians, attitudes towards homosexuality, and attitudes towards same-sex marriage.[119] We examined whether the relationship between legal regime and sexual orientation changes once controlling for these variables. We find that

---

[117] These are all statistically significant differences ($p < 0.01$ for the difference between religious agencies and the other agencies; $p < 0.1$ for the difference between public and private non-religious agencies).

[118] *See, e.g.*, Barak-Corren, *supra* note 97, at 104–05 (finding from a field experiment that *Masterpiece Cakeshop* increased discrimination towards same-sex couples in the wedding services market). In another study, the authors originally had not looked at the religiosity of the agency, but recoded their old data to examine whether religious agencies discriminated more than other agencies; their findings suggested that the answer is positive, but their study was underpowered. *See* Mackenzie-Liu, Schwegman & Lopoo, *supra* note 97, at 2, 8–10 (reanalyzing an audit study of foster care agencies that was two years prior to *Fulton* and finding discrepancies between same-sex and opposite-sex couples).

[119] Data on attitude towards homosexuality and same-sex marriage, and about the importance of religion and political tendency, was imported from *Religious Landscape Study*, Pew Rsch. Ctr., https://www.pewresearch.org/religion/religious-landscape-study [https://perma.cc/87AM-X8GC]. Data on religious affiliation was imported from *U.S. Religion Census Shows Both Stability and Change in Congregational Life*, U.S. Religion Census, https://www.usreligioncensus.org/node/1641 [https://perma.cc/3UDL-3W6N]. Data on states' GDP per capita was imported from *Gross Domestic Product by County, 2021*, U.S. Bureau of Econ. Analysis (Dec. 8, 2022, 8:30 AM), https://www.bea.gov/news/2022/gross-domestic-product-county-2021 [https://perma.cc/ZED7-M4K9].

the differences between jurisdictions with respect to sexual orientation disparities persist even when controlling for these socio-demographic and attitudinal variables.[120] That is, the association between law and the behavior of foster care agencies appears to exist independent of various attitudinal and demographic norms, providing additional weight to this finding.

In addition, it is important to remember that email inquiries do not capture the entire fostering process: We examine only the first stage of the licensing process. Disparate treatment can arise in many stages down the line, including in the interview phase, in home studies, and in the final placement of children in families. We do not test these additional stages and hence can offer no evidence on the total or further consequences of discrimination. Naturally, disparities that arise in the first contact phase reduce the number of options available for couples further down the line. To the extent that the process of contacting agencies has even moderate friction, one would expect that reduced positive responses to emails would ultimately translate into fewer opportunities for same-sex couples.

Ultimately, we are interested not only in the association between state laws and discrimination towards same-sex couples, but also in the causal effect of changes in the law: Did *Fulton* change the behavior of foster care agencies that were previously prohibited from discriminating? We examine this question now.

2. *Did the* Fulton *Decision Increase Discrimination Towards Same-Sex Couples?*

Post-*Fulton*, positive response rates were about 5.6 percentage points lower for all couples,[121] but the differences were bigger for same-sex couples. Yet what we found was a nuanced story: big differences

---

[120] Specifically, the significant interaction between sexual orientation and antidiscrimination law, showing that same-sex couples receive more positive responses in Antidiscrimination Only states, maintains its statistical significance and size of the coefficient, and same for the triple interaction between sexual orientation, antidiscrimination law, and religious exemption laws, showing that when both laws are present same-sex couples receive less positive responses. *See infra* app. C tbl.A3.

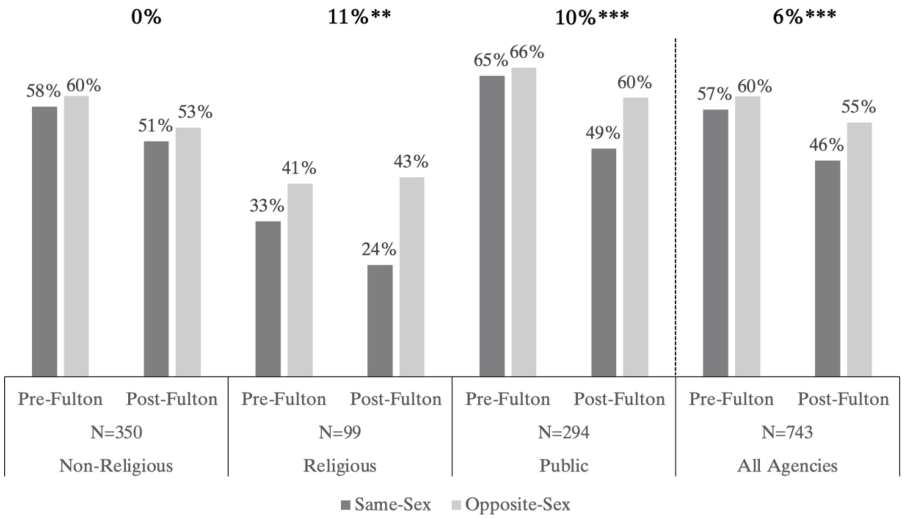
[121] The internal control group—the agencies kept aside and contacted for the first time only after *Fulton* was handed down—showed the same levels of positive response rates, indicating that the reduction in response was unrelated to previous waves. Notably the post-*Fulton* data is based on one wave, the third wave, in which agencies were randomly contacted by a same-sex couple or an opposite-sex couple. Thus, each sexual orientation subgroup includes approximately half of the agencies in the data. See the Appendix for details on exclusions. *See infra* app. I–J.

in some states, but not others.[122] On the whole, the results suggest that stakeholders understood the decision as a clear signal about the primacy of religious objections over antidiscrimination considerations, despite the seeming narrowness of the holding. Concomitantly, we found differences in response patterns between jurisdictions and types of agencies that suggest that the effect of *Fulton* was far from uniform.

a. The Effect of *Fulton* on Antidiscrimination Only Jurisdictions

In theory, *Fulton* was best positioned to make a direct legal impact in jurisdictions like Philadelphia (Antidiscrimination Only states).[123] Our results show that the sexual orientation gap in these states tripled post-*Fulton*, from 3 percentage points to 9 percentage points (see Figure 3).

FIGURE 3. CHANGES IN POSITIVE RESPONSE RATES BETWEEN SAME-SEX AND OPPOSITE-SEX COUPLES IN ANTIDISCRIMINATION ONLY JURISDICTIONS, BEFORE AND AFTER *FULTON*



* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The floating percentages represent the difference-in-differences from pre- to post-*Fulton*.

122 There were no order effects—namely, agencies that randomly received inquiries in the order of gay-straight-gay and agencies that received inquiries in the order of straight-gay-gay did not differ in their responses, post-*Fulton*.

123 These are the only jurisdictions where the decision was legally relevant: In jurisdictions that never enacted antidiscrimination rules, there was no binding norm to begin with, and in jurisdictions that enacted religious refusal laws, agencies were already allowed to discriminate.

This increase came from two sources: *religious agencies*, which developed a 19 percentage point gap between couple types, and *public agencies*, which developed an 11 percentage point gap between couple types, up from 4 percentage points. In contrast, private non-religious agencies maintained their preexisting gap of 2 percentage points between couple types.

b. Changes Among the Three Other Legal Regimes

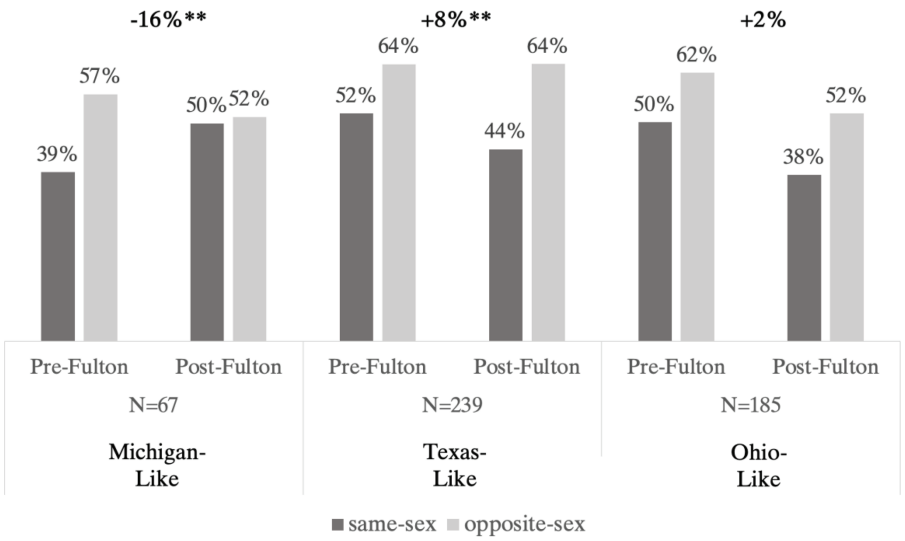
While agencies in these regimes were not subject to a relevant antidiscrimination obligation, or were exempt from such law on religious grounds, *Fulton* might have still influenced these jurisdictions. Perhaps, for example, religious agencies mistakenly believed they are subject to antidiscrimination laws. In such cases, *Fulton* could have updated these agencies' perceptions of their eligibility for exemption. Or, news about *Fulton* might have spilled over to *non-religious* (private or public) agencies—despite the irrelevance of the decision for these organizations—and as a result, agency employees who already harbored discriminatory preferences felt able to express them.

Figure 4 presents the overall results for the three regimes. Discrimination increased post-*Fulton* in the Texas-like jurisdictions (Religious Exemptions Only, n = 239). Agencies in these jurisdictions showed a 12 percentage point gap in positive responses between same-sex and opposite-sex couples before the decision, which increased to 20 percentage points after *Fulton* ($p < 0.01$). There was no statistically significant effect of *Fulton* on the positive response gap in Ohio-like jurisdictions (that enacted neither type of law, n = 185 agencies), where the gap between couple types increased from 12 percentage points to 14 percentage points.

Interestingly, agencies in jurisdictions that enacted both antidiscrimination and religious exemption laws (n = 67) showed a significant *decrease* in discrimination post-*Fulton*. Prior to *Fulton*, agencies in these jurisdictions showed the highest disparities in positive responses between same-sex and opposite-sex couples (18 percentage points). This gap was reduced to 2 percentage points after the decision ($p < 0.01$). This is a surprising result, as both states with relevant antidiscrimination laws (but no religious exemptions) and states with relevant religious exemption laws (but no antidiscrimination) separately experienced increased discrimination post-*Fulton*. This may be an artifact of a small sample. This group comprises only 3 states and 67 agencies, most of which are from Michigan (as compared with 185–743 agencies in each of the other jurisdiction types).[124]

[124] Notably, Michigan was implicated in a long and convoluted litigation between foster care agencies, same-sex couples, and the state at the time *Fulton* was decided, with two

FIGURE 4. CHANGES IN POSITIVE RESPONSE RATES BETWEEN SAME-SEX AND OPPOSITE-SEX COUPLES IN THE THREE NON-PERTINENT JURISDICTION TYPES BEFORE AND AFTER *FULTON*



* p < 0.1, ** p < 0.05, *** p < 0.01. The floating percentages represent the difference-in-differences from pre- to post-*Fulton*.

3. *Converging Evidence from a Longitudinal Survey*

In parallel to the field experiment involving foster care agencies, we conducted a companion longitudinal survey of the American public (n = 920)[125] which measured participants' support of service refusal by

open conflicts pending a decision in *Fulton*. *See infra* notes 129–32 and accompanying text. This setting might have been associated with a mixture of legal uncertainty and heightened social conflict that might be more complicated to analyze and deserves more in-depth investigation. For now, we will merely note that this result does resemble a pattern found in the Barak-Corren experiment around *Masterpiece Cakeshop*, a previous field experiment that examined discrimination towards same-sex couples in the wedding market before and after the *Masterpiece Cakeshop* decision. *See* Barak-Corren, *supra* note 97. That study found that regimes with both types of laws were the only regimes where discrimination towards same-sex couples did not increase.

125 We recruited participants from Prolific, an online participant recruitment platform that enabled us to recruit an ideologically and demographically representative sample with the appropriate quotas. We note that web-based samples lean younger, more urban, and slightly less conservative. *See* Scott Keeter & Kyley McGeeney, *Coverage Error in Internet Surveys*, PEW RSCH. CTR. (Sept. 22, 2015), https://www.pewresearch.org/methods/2015/09/22/coverage-error-in-internet-surveys [https://perma.cc/6RPE-J32Q]. Our balanced recruitment method aimed to ameliorate these tendencies. Recruiting returning participants is an additional challenge, which we dealt with first by collecting a very large sample—allowing attrition while

religious agencies to same-sex couples, before and after *Fulton*. Our goal was to learn whether *Fulton* changed public opinion on religious refusals. Participants were asked the following question both before and after *Fulton*: "To what extent do you oppose or support allowing foster care and adoption organizations with traditional beliefs about marriage to deny service to gay couples?" and provided their response on a 1–9 scale that ranged from strongly oppose to strongly support.

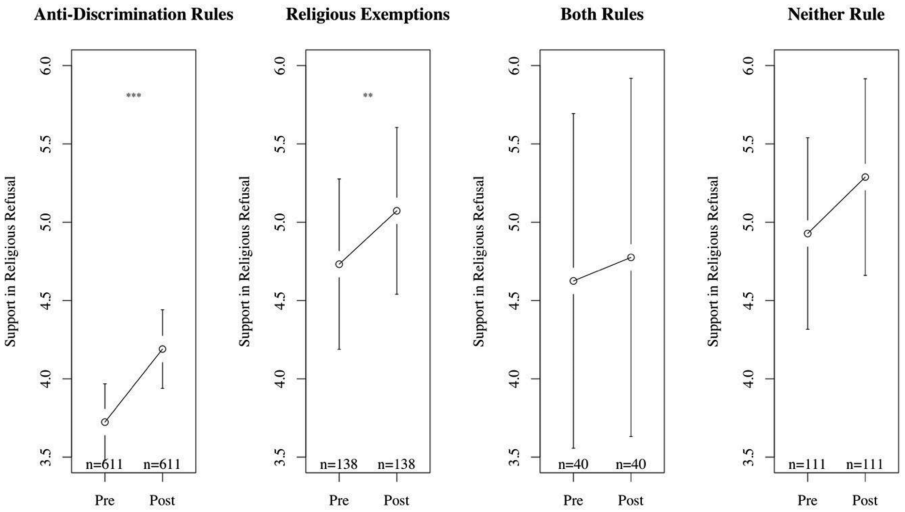
Our survey collected a host of demographics, including where participants reside, which we coded into legal regimes. The proportional representation of participants from the four legal regimes in the survey sample tracks closely the proportion of agencies from each regime in the field, contributing to the external validity of the survey. Notably, the two populations—members of the public and agency workers—are not identical, and likely differ on multiple grounds. Our results here provide additional triangulation of the evidence, exploring how *Fulton* influenced public opinion and attitudes on the question of foster care agencies' refusal.[126]

Our main finding is that the survey results correspond to the results of the field experiment. *Fulton* increased people's support for religious service refusal, in general and particularly in Antidiscrimination Only and Religious Exemption Only states—as was the case in the field experiment. Figure 5 plots these results. As it shows, survey participants became more supportive of religious service refusal post-*Fulton* in all four legal regimes.

preserving power, with 906 completed responses—and by monitoring the characteristics of retained versus attritted participants. *See* Robert M. Groves, Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer & Roger Tourangeau, Survey Methodology 379 (2004). The retention rate was high at 70% and we did not find that participants who attritted differed significantly than those who remained, demographically or ideologically, though we did find that earliest recruits were harder to retain.

[126] The survey is part of a separate project by the first author; here we describe the findings that are relevant for the present research questions. Its longitudinal design and *Fulton* predictions were pre-registered on aspredicted.org under pre-registration No. 68908. *Fulton Field Experiment (#64948)*, Wharton Credibility Lab, https://aspredicted.org/3ug6p.pdf [https://perma.cc/Y7F5-NX6R].

FIGURE 5. PUBLIC SUPPORT FOR ALLOWING RELIGIOUS FOSTER CARE AGENCIES TO REFUSE SERVICE TO SAME-SEX COUPLES BY LEGAL REGIME, BEFORE AND AFTER *FULTON* (N = 920)



* p < 0.1, ** p < 0.05, *** p < 0.01. Error bars represent 95% confidence intervals.

The proportional representation of participants from the four legal regimes in the survey sample tracks closely the proportion of agencies from each regime in the field, contributing to the external validity of the survey. There are notable differences between the four regimes that generally align with the findings from the field. First, in parallel to the behavior of the foster care agencies, survey participants from Antidiscrimination Only regimes that enacted no religious exemptions (left panel) had the lowest levels of support for religious service refusal, whereas participants in the other three regimes exhibited significantly higher levels of support.[127] Second, support for religious service

[127] Prior to *Fulton*, the average values of support for religious service refusal varied between the different legal regimes. On a scale on 1 to 9 (1 = strongly oppose; 9 = strongly support), the average among participants from Antidiscrimination Only regimes was 3.72, while the average values among participants from the other three regimes were higher (Religious Exemptions Only regimes = 4.73; Neither Rule regimes = 4.93; Both Rules regimes = 4.625). ANOVA testing showed statistically significant average differences between Antidiscrimination Only regimes and Religious Exemptions Only regimes (p = .004) and between Antidiscrimination Only regimes and Neither Rule regimes (p = .001). The difference between Antidiscrimination Only regimes and Both Rules regimes was not statistically significant (p = .291), likely because the small number of participants in this category (n = 40) produced a very noisy measurement (as the confidence interval for the mean of that category shows). *See supra* fig. 5.

refusal rose after *Fulton*, with the most significant effects showing in the Antidiscrimination Only and Religious Exemptions Only regimes, also paralleling the results from the field. The attitudes of participants in Michigan-like regimes with both antidiscrimination and religious exemption laws change more modestly and the change is not statistically significant. In any case, we do not observe any compensatory trend in this small category of survey participants. This is the only legal regime where the survey results do not track the field results, but in both cases the size of the group is very small. This discrepancy yields the conclusion that it might not be advisable to rely too much on the field results with respect to this small and potentially non-generalizable group.

C. Discussion

There are two stories in these results. The first is that the Supreme Court's ruling made a difference in the one place we would expect it to: it taught agencies in jurisdictions that previously lacked religious exemptions that they could behave more like agencies in jurisdictions that had them. Foster agencies in Minnesota (as well as in Kentucky, California, and Idaho, for example) were free to discriminate. And they chose to, at least in how they greeted potential clients knocking on their door.

But, as we found, the data tells a second story: *Fulton* appeared to license even more discrimination in places where it formally did not control, and by agencies which it did not nominally speak to. Thus, non-religious and public agencies in Texas-like states with only religious exemption laws more than doubled their sexual orientation disparities, from 1% to 17% and from 11% to 23%, respectively.[128] But under no reading of *Fulton* are these non-religious agencies now entitled to *religious* exemptions.

Relatedly, increases in sexual orientation disparities in the non-legally pertinent regimes mostly did not stem from increased discrimination from religious agencies. In one jurisdiction type (religious exemption-only), religious agencies maintained their prior levels of positive responses to same-sex couples (Figure A1). In another (antidiscrimination and religious exemption), religious agencies responded more favorably to same-sex

[128] We cannot relay any findings with respect to Michigan-like jurisdictions due to the very small number of public agencies (n = 3) in this jurisdiction type. Our findings regarding Ohio-like jurisdictions are positive with respect to public agencies, whose differential, negative treatment of same-sex couples doubled (from 8% to 16%), but not statistically significant. This result may be underpowered.

couples post-*Fulton* (Figure A2).[129] Only in Ohio-like jurisdictions, with neither antidiscrimination nor religious exemption rules, did disparities increase among religious agencies, but this result was not statistically significant (Figure A3). Overall, religious agencies appear to harbor no mistaken beliefs about their entitlement to religious exemptions.

We can only speculate as to what happened as the Court's ruling made its way into the agencies' ears. Employees of these agencies might have interpreted *Fulton* as generally supportive of religious objections to same-sex parenting, inferring that they may be available for individual objectors, or that *Fulton* shaped their preferences about working with same-sex couples. It is hard to assess which option is more likely. Presumably, individuals who hold religious objections to same-sex couples and work in a position that might bring them to encounter same-sex couples would *know* whether they are subject to laws that compel them to act against their beliefs—thus *Fulton* should not have had an impact on them. However, *Fulton* was still legal news, and as such could have led to inferences about new refusal rights. At the same time, it is also possible that the decision tipped the scale for agents who were "on the fence" about serving same-sex couples—not for legal reasons, but for the normative statement it expressed about the primacy of religious liberty. For these agents, it would be more plausible to assert that *Fulton* changed individual preferences rather than freed them to act on their existing preferences.

Our results open up a series of intriguing questions. As states appear to differ substantially in their patterns of discrimination, the relative strength of compelling interest arguments may differ from one state to another. For example, states that show *more* favorable treatment of same-sex couples than opposite-sex couples may have a harder case claiming that they have a compelling interest in eradicating sexual orientation discrimination, as this discrimination has (empirically speaking) been eradicated and even reversed, at least in this domain. In contrast, the compelling interest of states with substantial rates of discrimination can be vindicated by the findings. Our Appendix contains descriptive statistics for all fifty states, presenting their legal regimes and average rates of discrimination as documented throughout this study. We note again that state-level results should be interpreted with caution due to the sometimes small number of agencies per state.

[129] This behavior appears compensatory, but it remains hard to know why agencies in Michigan-like regimes, of all places, responded to *Fulton* in this way. We note again that this is a small group of agencies, mostly from one particular state immersed in legal conflict, and therefore these results could be less generalizable. In our companion study that included longitudinal surveys measuring public attitudes before and after *Fulton*, we found no similar effect in Michigan-like jurisdictions.

The larger the number of agencies, the more precise are the estimates. Figures beginning with A can be found in the Appendix.

III
Evidence from Litigation: How Parties Thought *Fulton* Applied

The *Fulton* opinion focused on very specific contractual language and decisively avoided setting a broader framework for the conflict at the heart of the dispute. As Justice Alito noted in concurrence, this choice made it legally possible for Philadelphia to simply change its contract and continue to deny an exemption to the refusing Catholic agency.[130] Justice Gorsuch explained that "with a flick of a pen, municipal lawyers may rewrite the City's contract to close the . . . loophole."[131] The evidence from our nationwide study of foster care agencies suggests that the decision did not dissolve on contact as Justice Alito feared. Foster care agencies in the immediate aftermath of the decision became less willing to provide service to same-sex couples.[132]

This Part adds another layer of data by surveying the outcomes of all conflicts between equality-seeking governments and objecting religious foster care agencies that were litigated post-*Fulton*. Our survey shows that courts and litigants also read the opinion broadly. The predictions of Justices Gorsuch and Alito were not borne out by legal reality.

The first case to observe is *Fulton* itself. Despite Justice Gorsuch's concern that "municipal lawyers may rewrite the City's contract" to circumvent the holding, Philadelphia and CSS settled after the *Fulton* decision,[133] and Philadelphia agreed to pay CSS two million dollars and to renew the contract without requiring that the agency comply with the city's nondiscrimination policy.[134] Although the case was remanded, "the city chose not to continue the fight."[135] According to Justice Alito's concurrence, Philadelphia could have just removed its never-used exemption clause from the contract in order to prohibit the agency from discriminating. That the parties chose to settle suggests that they may have viewed the decision as a more decisive position on the conflict.

---

130 *See* Fulton v. City of Philadelphia, 141 S. Ct. 1868, 1887 (2021) (Alito, J., concurring) ("[I]f the City wants to get around today's decision, it can simply eliminate the never-used exemption power.").

131 *Id.* at 1930 (Gorsuch, J., concurring).

132 *See supra* Part II.

133 *See* Joint Motion for Consent Judgment at 2–5, Fulton v. City of Philadelphia, 320 F. Supp. 3d 661 (E.D. Pa. 2021) (No. 18-2075) (describing the settlement between the parties).

134 *Id.*; *see also* Brownworth, *supra* note 37.

135 Brownworth, *supra* note 37.

This outcome was not unique. All similar disputes seem to have been negotiated after *Fulton* in the shadow of a presumption in favor of religious exemptions. In Michigan, a conflict between St. Vincent Catholic Charities and the state that was pending when *Fulton* was decided[136] was settled on the grounds that the agency would likely prevail as a consequence of *Fulton*.[137] A second conflict, wherein Catholic Charites West Michigan alleged that the Michigan Department of Health and Human Service's policy of prohibiting discrimination against same-sex couples violated its constitutional rights,[138] was resolved in an identical fashion. In a stipulation resolving all claims, the parties noted that in light of *Fulton*, the foster care agency "would likely prevail" on its free exercise claim and agreed to enter judgment against the state.[139] This outcome unfolded even though, unlike *Fulton*, the contract between the agencies and the state of Michigan included no language similar to the never-used discretionary exemption that the Court viewed as dispositive in *Fulton*. *Fulton*'s main holding could have been irrelevant, but the parties and the court still behaved as though bound by a new rule in favor of religious exemptions.

In Kentucky, a months-long conflict between the state and a Baptist foster care agency was resolved shortly after *Fulton*.[140] At the center of the conflict was a contract that banned LGBTQ discrimination by private agencies that provide foster and adoptive services. Sunrise Children's Services refused to sign the contract, arguing that it violated its religious beliefs.[141] Notably, the contract had no similar language to the Philadelphia contract. Rather than insist on its enforcement or inoculate it further against the *Fulton* holding, Kentucky resolved the conflict by dropping the terms "sexual orientation" and "gender identity" from the antidiscrimination clause of the contract (which also

136 *See generally* Complaint, Buck v. Gordon, 429 F. Supp. 3d 447 (W.D. Mich. 2019) (No. 19-cv-00286).

137 Stipulated Order & Judgment Resolving All Claims Against State Defendants at 4, Buck v. Hertel, No. 1:19-CV-00286 (W.D. Mich. Jan. 26, 2022).

138 *See* Verified Complaint for Declaratory & Injunctive Relief at 25–26, Cath. Charities W. Mich. v. Mich. Dep't of Health and Hum. Servs., No. 19-000072-MM (Mich. Ct. Cl. Apr. 24, 2019).

139 Stipulation Resolving All Claims at 4, Cath. Charities W. Mich. v. Mich. Dept. of Health and Hum. Servs., No. 19-CV-11611 (E.D. Mich. Mar. 21, 2022), https://adfmedialegalfiles.blob.core.windows.net/files/CatholicCharitiesWMStipulationAndOrder.pdf [https://perma.cc/8RUW-S76M].

140 *See* Deborah Yetter, *Kentucky Resolves Contract Feud with Baptist Children's Agency over LGBTQ Discrimination*, Louisville Courier J. (July 16, 2021), https://www.courier-journal.com/story/news/politics/2021/07/16/kentucky-contracts-baptist-sunrise-childrens-services-after-lgbtq-feud/7989280002 [https://perma.cc/6JFN-SEY6].

141 *Id.*

banned discrimination because of race, religion, sex, age, or disability).[142] The revised contract requires Sunrise to refer LGBTQ applicants to another agency.[143] The parties agreed that *Fulton* controlled the result and Governor Beshear, a Democrat, explained that "while not directly on point, [*Fulton*] is pretty close."[144] Here too, *Fulton*'s main holding could have been irrelevant, yet the parties thought that the decision controlled their conflict.

A pair of lawsuits in New York were also resolved in favor of the foster care agency in the fall of 2022. In the first, *New Hope Family Services, Inc. v. Poole*, a religious foster care agency sued the state of New York over the state's insistence that the agency either revise its policy of noncompliance with the state's nondiscrimination law or terminate its adoption program.[145] The court subjected the state's nondiscrimination law to strict scrutiny and found that, while the state's interests were compelling, they were not narrowly tailored.[146] The parties ultimately settled, and the foster care agency was allowed to refuse service to same-sex couples and continue its work.[147] In a second lawsuit involving the same agency, the court explained that *Fulton* applied "persuasively" and concluded that the agency was not a place of public accommodation bound by the state's antidiscrimination law.[148] Here, too, *Fulton* was interpreted as a broad decision that controls the outcome of cases with very different circumstances.

Most recently, two twin federal lawsuits ended in the same fashion, confirming South Carolina's position that it should not be denied federal monies for foster care and adoption programs because it exempts agencies from antidiscrimination rules.[149] That each of these conflicts was

---

142 *Id.*

143 *Id.*

144 *Id.*

145 New Hope Fam. Servs., Inc. v. Poole, 626 F. Supp. 3d 575, 578–79 (N.D.N.Y. 2022).

146 *Id.* at 584–85. In a prior phase of litigation between these parties, the Second Circuit observed that "New Hope's pleadings easily give rise to the 'slight suspicion' of animosity," New Hope Fam. Servs., Inc. v. Poole, 966 F.3d 145, 165 (2d Cir. 2020), a comment that led the district court on an earlier remand to conclude that the law was not neutral and based on religious hostility, New Hope Fam. Servs., Inc., v. Poole, 493 F. Supp. 3d 44, 59 (N.D.N.Y. 2020). Thus, in its 2022 decision, the Court did not need to reference *Fulton* based on this earlier finding. *New Hope Fam. Servs., Inc.*, 626 F. Supp. 3d at 584–86.

147 New York paid the plaintiffs $250,000. *See* New Hope Family Services v. Poole, ALL. DEF'G FREEDOM (Mar. 7, 2023), https://adfmedia.org/case/new-hope-family-services-v-poole [https://perma.cc/EDF2-5MV4] (detailing the terms of the settlement agreement between New Hope Family Services and the state of New York); Settlement Agreement at 2–3, *New Hope Fam. Servs., Inc.*, 626 F. Supp. 3d 575 (No. 18-cv-1419).

148 New Hope Fam. Servs., Inc., v. James, No. 5:21-CV-01031, 2022 WL 4494277, at *5, *40 (N.D.N.Y. Sept. 28, 2022).

149 *See* Rogers v. McMaster, No. 6:19-cv-01567, 2023 WL 7396203, at *1 (D.S.C. Sept. 29, 2023) (granting summary judgment in favor of the defendants in a case where plaintiffs

resolved in favor of religious agencies in light of the Court's apparently fact-specific decision in *Fulton* suggests that the *Fulton* holding was widely understood as a clear statement about the entire category of religious agency-nondiscrimination conflicts.

Seeking to understand this pattern, we conducted background conversations with some of the lawyers who were involved in settling these conflicts.[150] What they shared was a similar narrative. "Neither me or any of the parties took the opinion as law," said one lawyer. "The Chief Justice entirely twisted Pennsylvania law to get at this outcome, he was trying to make the case go away. Nobody believed the majority believed its own opinion."

What, then, did lawyers understand the Court to do? "What they're really doing is sending a message. If it sees that [we] are willing to defy it, nobody thought the Court is going to blink and change course," said one of the lawyers who had to decide whether to continue litigation or settle the case. He further stated:

> "[It's as if the Court was saying],'we are going to reverse what happened here and do almost no law while doing so, and send a warning shot.' The very clear implication was that we want you to figure a way to

challenged South Carolina's accommodating a religious agency that wanted to work only with "Christian foster parents who can affirm its doctrinal statement of faith," in part because *Fulton* controlled the outcome); Maddonna v. U.S. Dep't of Health & Hum. Servs., No. 6:19-CV-3551, 2023 WL 7395911, at *1 (D.S.C. Sept. 29, 2023) (same); *see also Rogers v. Health and Human Services*, ACLU (July 19, 2023), https://www.aclu.org/cases/rogers-v-health-and-human-services [https://perma.cc/8J8M-S4JQ] (explaining the lawsuit alleging that the federal Department of Health and Human Services' (HHS's) granting religious exemptions for foster care agencies in 2019 violates the Establishment Clause). These cases stem from nondiscrimination rules for foster care agencies that were issued by HHS, though HHS later declared that it would not enforce the rules. A few cases were dismissed on these grounds, prior to the 2023 rulings. *See, e.g.*, Texas v. Azar, 476 F. Supp. 3d 570, 580 (S.D. Tex. 2020); Holston United Methodist Home for Child., Inc. v. Becerra, No. 2:21-cv-185, 2022 WL 17084226, at *21 (E.D. Tenn. Nov. 18, 2022). There is also a line of cases that challenges HHS nonenforcement policy. *See, e.g.*, Corrected Complaint for Declaratory and Injunctive Relief, Fam. Equal. v. Azar, No. 1:20-cv-02403 (S.D.N.Y. Apr. 1, 2020), https://lambdalegal.org/case/family-equality-v-azar [https://perma.cc/GL6V-QSVR] (announcing the commencement of litigation against the Department of Health and Human Services for their nonenforcement policy); Complaint, Facing Foster Care in Alaska v. U.S. Dep't of Health & Hum. Servs., No. 1:21-cv-00308 (D.D.C. Feb. 2, 2021). As noted, some of these cases resulted in wins for religious refusals, others were dismissed as dead disputes, and a few others are still pending at this time. No case since *Fulton* has resulted in a holding that HHS, or any state, is entitled to enforce its antidiscrimination policy without providing religious exemptions.

[150] These conversations were conducted under conditions of full anonymity and without interference with attorney-client privilege, so the lawyers are unidentified. The lawyers were not asked about the parties or any case-specific advice, but about their legal interpretations of the *Fulton* decision. All interviews were conducted by phone between July 6, 2023 and July 10, 2023.

> solve it in the interest of the refusing agencies, and if not we're going to do it, we'll do so ourselves with much broader implications."

The overwhelmingly uniform pattern of outcomes in post-*Fulton* litigation, coupled with the lawyers' view of the decision as a clear message with clear implications for foster care agencies, both help explain the behavior of foster care agencies post-*Fulton.* Together, the behavioral field evidence and the litigation evidence show that *Fulton* had a substantial and broad effect, both in its immediate aftermath and in the longer term.

IV
Consequential Constitutional Law

"The first call of a theory of law is that it should fit the facts" wrote Oliver Wendell Holmes in 1881.[151] Over a hundred years later, in 1998, Richard Posner wrote *Against Constitutional Theory*, sharply criticizing the Supreme Court for making decisions in constitutional cases "so barren of any engagement with reality that the issue of their correctness scarcely arises" and blaming constitutional theory for claiming "to offer the courts a data-free method of deciding cases, rather than helping in the discovery and analysis of the relevant data."[152] Posner argued that "it is the lack of an empirical footing that is and always has been the Achilles heel of constitutional law, not the lack of a good constitutional theory. But this raises the question of what the courts are to do in difficult constitutional cases when their ignorance is irremediable."[153]

Two decades have passed since then. Constitutional law still largely lacks an empirical footing. Courts still decide cases with little attempt to discover and analyze relevant data.[154] And constitutional debates still miss good evidence of how Supreme Court rulings affect individual behavior.[155] The evidence we provided in this article shows that this

---

151 Oliver Wendell Holmes, The Common Law 211 (1881).

152 Posner, *supra* note 1, at 18, 21.

153 *Id.* at 21.

154 Indeed, the current Chief Justice expressed outright hostility to the idea. *See* Transcript of Oral Argument at 39–40, Gill v. Whitford, 138 S. Ct. 1916 (2017) (No. 16-1161) (Roberts, C.J.) (dismissing an attempt to quantify gerrymandering as "sociological gobbledygook"). However, as we show later in this Part, this approach is not necessarily representative or exhaustive of the Court's approach to evidence. As we explain, the Supreme Court has expressed a need for more evidence to substantiate compelling government interests, but lacks a coherent, consistent method of applying it.

155 *See* James E. Ryan, *The Limited Influence of Social Science Evidence in Modern Desegregation Cases*, 81 N.C. L. Rev. 1659, 1664 (2003) (noting that despite its potential usefulness, social science evidence probably has "little influence on modern desegregation decisions"); Ryan D. Enos, Anthony Fowler & Christopher S. Havasy, *The Negative Effect*

omission is misguided, as it undermines the ability of constitutional law to meet its own goals and prevents the Court from understanding how its decisions operate in the real world.

This Part sets out some preliminary gestures towards a new kind of pragmatic constitutional theory—consequential constitutional law. We begin in Section IV.A by situating our work within the larger context of scholarly work examining constitutional consequences.[156] Our findings converge with previous scholarly work on the effects of Supreme Court decisions and religious exemption laws, expanding the evidentiary basis for consequential constitutional law. Section IV.B then adds legitimacy consequences of the Supreme Court's incremental approach to the discussion, pointing to the potential interplay between different kinds of consequences that constitutional law cares about. Finally, in Section IV.C, we demonstrate the current empirical deficiencies in free exercise doctrine and show how our approach can remedy these deficiencies and make the application of constitutional doctrine to specific cases more grounded and nuanced.

A. Converging Evidence on Social Consequences of the Law

Research on the Supreme Court's effect on public opinion goes back decades.[157] We will focus here on the effects of the Court's religious exemption jurisprudence on sexual orientation discrimination and acceptance, and, more generally, on the effects of the Court's decisions on public opinion.

Fallacy: A Case Study of Incorrect Statistical Reasoning by Federal Courts, 14 J. Emprical Legal Stud. 618, 625–29 (2017) (describing a series of cases in which a Supreme Court majority opinion fell prey to "the negative effect fallacy," a statistical error); *id.* at 637–38 (discussing why the Supreme Court may employ the negative effect fallacy).

156 We focus this Section on the context of this article, namely religion/equality conflicts, yet consequences are relevant for a wide scope of constitutional questions. *See, e.g.*, Roseanna Sommers & Vanessa K. Bohns, *The Voluntariness of Voluntary Consent: Consent Searches and the Psychology of Compliance*, 128 Yale L.J. 1962, 2011–19 (2019) (finding that assumptions on which search and seizure doctrine rest are not borne out by empirical studies). There is definitely a need for an overarching empirical framework in constitutional law, a task we leave for future work.

157 *See, e.g.*, Robert H. Birkby, *The Supreme Court and the Bible Belt: Tennessee Reaction to the "*Schempp*" Decision*, 10 Midwest J. Pol. Sci. 304, 307–11 (1966) (describing an empirical study of the effects of the *School District of Abington Township v. Schempp* decision, which struck down mandatory Bible reading in public schools as a violation of the Establishment Clause, on school boards in Tennessee); Stephen L. Wasby, *The Supreme Court's Impact: Some Problems of Conceptualization and Measurement*, 5 L. & Soc'y Rev. 41, 45 (1970) (reviewing empirical scholarship of Supreme Court decisions to that date); David G. Barnum, *The Supreme Court and Public Opinion: Judicial Decision Making in the Post-New Deal Period*, 47 J. Pol. 652, 654–61 (1985) (conducting an empirical study of public reaction to Supreme Court decisions).

1. *Effects on Service Providers*

Our findings on the social and legal effects of the *Fulton* decision on foster care agencies are in line with the results of prior studies, most relevantly a similar field experiment that was conducted around the Court's decision in *Masterpiece Cakeshop*.[158] The *Masterpiece Cakeshop* experiment examined the behavior of wedding vendors in four jurisdictions of the same types as in the present study, before and after the decision.[159] Similar to the present study, the *Masterpiece Cakeshop* field experiment found that wedding vendors showed a "robust reduction in the willingness to serve same-sex couples, as compared with opposite-sex couples" following the decision's release.[160] This reduction occurred even in jurisdictions where the decision did not modify the law and no antidiscrimination rule ever applied to the vendors.[161]

Together, these two studies offer data points from different social contexts (for-profit businesses and non-for-profit public and private foster care agencies) and areas of law (public accommodations law and foster care law), in two different periods, and around different Supreme Court decisions. They show that sexual orientation discrimination is a lingering issue, that different jurisdictions are associated with discernibly different patterns of discrimination (with the results being consistent between the two studies with respect to each jurisdiction type), and that even a seemingly narrow judicial decision to grant a religious exemption has broad effects.

The two studies complement each other. First, the *Masterpiece Cakeshop* field experiment did not have access to businesses' religiosity and therefore did not have a direct identification of the role of religiosity in the documented effect.[162] The *Fulton* experiment, by coding publicly available information on agency religiosity, is able to identify and evaluate the role of religiosity. Thanks to this improvement, we are able to learn that religious agencies engage in sexual orientation discrimination more than other agencies. We learned that they appear to be fully aware of their refusal rights and used them where they were available before *Fulton*. Finally, we learned that where previously unavailable exemptions opened up post-*Fulton*, religious agencies altered their behavior accordingly.

---

158 *See* Barak-Corren, *supra* note 97.

159 *Id.* at 77, 84–86. That project involved more limited data collection focused on sampling businesses from four states that resembled each other in all pertinent aspects except their legal domain. The *Fulton* experiment is a nationwide population study that encompasses the vast majority of foster care agencies in the United States.

160 *Id.* at 104.

161 *Id.* at 106.

162 *Id.* at 101–02.

At the same time, we see that religious agencies are not the only type of agencies attentive to the Supreme Court. Non-religious organizations become less likely to respond positively to same-sex couples as well, post-*Fulton*.[163] The introduction of a new constitutionally mandated *religious* exemption has sent ripples to *non-religious* organizations. This illuminates an issue that was left open after the *Masterpiece Cakeshop* study.

2. Effects on LGBTQ Individuals

How does increased discrimination by service providers influence LGBTQ people? Neither the *Masterpiece Cakeshop* nor the *Fulton* experiments measured these effects directly as both studies examined the behavior of service providers through the audit procedure. Yet in recent years, researchers have found that when discrimination against LGBTQ people is declared legal, it has psychological and social consequences for the targets of discrimination.

In one study, Professor Julia Raifman and her colleagues studied mental health data from 2014 through 2016 in three states that implemented laws permitting the denial of services to same-sex couples at that time (Utah, Michigan, and North Carolina) and six nearby control states (Idaho and Nevada, Ohio and Indiana, and Virginia and Delaware, respectively).[164] The researchers found that the proportion of sexual minority adults reporting mental distress increased from 21.9% to 32.1% during that period in states that passed denial-of-service laws as compared with control states—a 46% relative increase.[165] Laws permitting denial of services to same-sex couples were not associated with significant changes in heterosexual adults experiencing mental distress.[166] Notably, these results respond to legislature-enacted exemptions rather than Court-mandated exemptions, so some differences probably apply.[167] In both settings, however, the legal change exposes LGBTQ people to more service refusals or places them under the increased possibility that they could experience such refusals.[168]

---

163 *See supra* Section III.B.2.

164 Julia Raifman, Ellen Moscoe, S. Bryn Austin, Mark L. Hatzenbuehler & Sandro Galea, *Association of State Laws Permitting Denial of Services to Same-Sex Couples with Mental Distress in Sexual Minority Adults: A Difference-in-Difference-in-Differences Analysis*, 7 JAMA Psychiatry 671, 672 (2018).

165 *Id.* The authors of this study defined participants as "sexual minorities" if they "identified as gay, lesbian, bisexual, or not sure of their sexual orientation." *Id.* at 672.

166 *Id.* at 674.

167 *Id.* at 676.

168 *Id.* at 672.

Raiffman and her colleagues proposed that this mechanism was the cause of the increased mental distress.[169]

Moving from religious exemptions to other Supreme Court decisions and their effects on LGBTQ individuals, previous studies found that the Court has shifted public opinion in various directions following its *Obergefell* decision to legalize same-sex marriage.[170] Two studies that measured the effect of *Obergefell* on public attitudes documented an increase in the perception that social norms support same-sex marriage,[171] and in personal support for same-sex marriages.[172] Another study, by Professor Eugene Ofosu and colleagues, found a sharper decrease in antigay bias in states that legalized same-sex marriage compared with those that did not.[173] However, for the fifteen states that did not pass same-sex marriage legalization locally by the time *Obergefell* was decided, antigay bias increased over time after the decision (notwithstanding an overall trend of bias reduction in the years before *Obergefell* in these states).[174] These findings cast *Obergefell* in a more complicated light than is commonly appreciated: the decision won an expansive constitutional right for LGBTQ communities, but at the cost of an inadvertent increase in LGBTQ bias. Understanding that law and high court decisions can have a multitude of effects, both desirable and inadvertent, is an important first step.

Overall, both the present Article and previous studies provide strong evidentiary basis for consequential constitutional law: The Court's decisions, as well as the laws they review and scrutinize, are social phenomena with cultural influence. These effects can be measured and analyzed, and—to the extent that consequences play a role in constitutional doctrine—we now have a rich toolkit of empirical

169 *Id.* at 675.

170 Obergefell v. Hodges, 576 U.S. 644 (2015).

171 Margaret E. Tankard & Elizabeth Levy Paluck, *The Effect of a Supreme Court Decision Regarding Gay Marriage on Social Norms and Personal Attitudes*, 28 PSYCH. SCI. 1334, 1337, 1339 (2017).

172 Emily Kazyak & Mathew Stange, *Backlash or a Positive Response?: Public Opinion of LGB Issues After* Obergefell v. Hodges, 65 J. HOMOSEXUALITY 2028, 2044–47 (2018).

173 Eugene K. Ofosu, Michelle K. Chambers, Jacqueline M. Chen & Eric Hehman, *Same-Sex Marriage Legalization Associated with Reduced Implicit and Explicit Antigay Bias*, 116 PROC. NAT'L ACAD. SCIS. 8846, 8846 (2019) (noting that even among states that did not legalize same-sex marriage locally, implicit and explicit antigay bias were declining in the years preceding federal legalization).

174 *Id.* at 8849–50. The results of the Ofosu, Chambers, Chen, and Hehman study point to the possibility that in the long term, the *Fulton* decision might cause anti-religious bias in states that were forced to introduce religious exemptions to their laws (the seeming equivalent of what *Obergefell* had been for states that did not legalize same-sex marriage prior to *Obergefell*, and where the decision seemingly increased antigay bias). This is a notable concern deserving of future research.

methods to study many of these consequences and understand the legal and social factors that drive them.

B. The Interplay Between Social Consequences and Legitimacy Consequences

The social consequences of Supreme Court decisions are not the only type of consequences that actors within the legal system—and particularly Supreme Court Justices—care about. In recent years, the Roberts Court arguably has used narrow decisions and coalition-building as one of several strategies to preserve the Supreme Court's institutional legitimacy.[175] In *Masterpiece Cakeshop*, for example, the Court managed to decide an LGBTQ rights versus religious freedom question 7-2, by avoiding the direct conflict and focusing instead on a procedural issue.[176] The narrow ruling in *Fulton* managed to secure unanimity among the Justices by avoiding the direct conflict between religion and equality, this time by focusing on contractual interpretation.[177] These deliberately narrow decisions—framed, in both cases, by sharp dissents and concurrences accusing the majority of running away from a decision appear to reflect a concern about a potential blow to the Court's institutional legitimacy.[178]

Justices repeatedly invoke the Court's public standing and perceived legitimacy in their opinions.[179] On several occasions, the Court has been criticized for (or perhaps accused of) of orchestrating hot-button decisions to align with public sentiment[180] or to avoid media

---

175 *See* Donnelly, *supra* note 13, at 1507–11 (describing several cases in which Chief Justice Roberts built coalitions and gave both the left and the right "a partial victory").

176 *See* Masterpiece Cakeshop, Ltd., v. Colo. C.R. Comm'n, 138 S. Ct. 1719, 1732 (2018) (deciding in favor of the appellant on the grounds he was denied adjudication by a "neutral decisionmaker who would give full and fair consideration to his religious objection").

177 Fulton v. City of Philadelphia, 141 S. Ct. 1868, 1878 (2021) (holding that a contractual non-discrimination requirement was rendered "not generally applicable" by a "formal system of entirely discretionary exceptions").

178 *See, e.g.*, *id.* at 1881 (Alito, J., concurring); *Masterpiece Cakeshop*, 138 S. Ct. at 1751–52 (Sotomayor, J., dissenting).

179 *See* Gillian E. Metzger, *Considering Legitimacy*, 18 Geo. J. L. & Pub. Pol'y 353, 375–76 (2020) (discussing several examples of opinions in which Justices make reference to public perception and related concerns).

180 *See, e.g.*, Tara Leigh Grove, *The Supreme Court's Legitimacy Dilemma*, 132 Harv. L. Rev. 2240, 2255 (2019) (reviewing Richard Fallon Jr., Law and Legitimacy in the Supreme Court (2018)) (noting that in "politically charged moments," legal scholars believe Justices changed positions to protect the Court); *id.* at 2269 (noting that vote-switching is part of a tradeoff between legal legitimacy and perceived legitimacy); *id.* at 2254 (describing reports that Chief Justice Roberts switched his vote to upholding the Affordable Care Act's individual mandate in *National Federation of Independent Businesses v. Sebelius* in response to "a barrage" of media criticism).

criticism.[181] Interestingly, Fulton was released to the public alongside *California v. Texas*,[182] a lopsided 7-2 decision that saved the Affordable Care Act from yet another challenge.[183] The simultaneous release of the two decisions gave victories to both liberals and conservatives, and may have been intended to portray the Court as a neutral institution. A few days after it released the *Fulton* decision, the Court decided *Mahanoy Area School District v. B.L.*, 8-1.[184] There, the Court held that a public school district violated a cheerleader's First Amendment rights when they disciplined her for posting a Snapchat video of herself expressing profane frustration for not making the varsity team.[185] In an ambiguous opinion, the majority explained it would "leave for future cases to decide" when a school could punish an off-campus speaker.[186] Professor Suk Gersen argued that this "uncategorical and vague" decision was "all the better for keeping the liberal-conservative coalition on board."[187]

1. The Consequential Challenges for Preserving Legitimacy

Whether the Court *should* weigh legitimacy consequences is a serious and difficult question. A separate question is whether the Court can actually *control* the legitimacy consequences of its decisions. One issue is that the Court is not in full control of how its decisions are subsequently framed and perceived by the public. Professors Michael Nelson and James Gibson argue that the public's perception of the Court's legitimacy depends on the Justices' perceived motivations.[188]

---

181 Whether deciding cases based on public opinion is a valid judicial act is another debate. For the opinion that the Court should not switch its vote to protect its public image, see *id.* at 2254–55, 2262 (arguing that such considerations violate consistency, justice, and reason). For the contrary argument that public opinion is a valid and important consideration, see Metzger, *supra* note 179, at 376–77 (arguing that these considerations are one reason among others and are critical to the Court's function and survival because the Court must ensure that those who lose abide by the decisions). Notably, judicial minimalism can be defended on other grounds as an exercise in modesty that supports deliberative democracy in controversial issue areas. *See generally* Cass R. Sunstein, One Case at a Time: Judicial Minimalism on the Supreme Court (1999) (praising judicial minimalism in cases concerning such topics as abortion, affirmative action, free speech, homosexuality, and sex discrimination).

182 California v. Texas, 141 S. Ct. 2104, 2120 (2021) (holding that the plaintiffs did not have standing to challenge the Affordable Care Act's minimum essential coverage provision).

183 For a critique of the Court's approach in the first challenge, see *supra* note 180 and accompanying text.

184 Justice Thomas was the only dissenter. Mahanoy Area Sch. Dist. v. B.L., 141 S. Ct. 2038, 2042 (2021).

185 *Id.* at 2043, 2048.

186 *Id.* at 2046.

187 Gersen, *supra* note 13.

188 James L. Gibson & Michael J. Nelson, *The Legitimacy of the US Supreme Court: Conventional Wisdoms and Recent Challenges Thereto*, 10 Ann. Rev. L. & Soc. Sci. 201, 210–12 (2014).

These perceptions depend on how an individual was exposed to the decision and what intermediary, if any, reported on the opinion.[189] These hypotheses are supported to some degree by the work of Professors Katerina Linos and Kimberly Twist, who showed that the framing of Supreme Court decisions in the media influences public perceptions and understanding of the decisions.[190] To a large extent, this factor may be outside of the Court's control.

A second challenge to the ability to maintain legitimacy involves the public tendency to react to the Court based on the bottom line outcomes of its decisions. In their empirical study, Professors Logan Strother and Shana Kusher Gadrian find that policy disagreement with the outcomes of Supreme Court cases is connected to the public's perception of the Court's legitimacy.[191] They find that policy disagreement with the outcome of a case leads to the perception that the Court is "political" and therefore less legitimate.[192] Notably, they find that legitimacy did not change for study participants who disagreed and agreed with the same number of decisions.[193] In their words, "policy wins and losses offset each other."[194]

These findings create an interesting question for cases like *Fulton*. If the public considers *Fulton* primarily alongside *California v. Texas*, the Court may be able to control potential damage to its public image by offsetting wins and losses for partisans. If, on the other hand, the public considers *Fulton* against the long series of cases where the Court favored religious interests consistently over other interests, including sexual orientation and gender equality, then ruling narrowly may not be an especially effective strategy to preserve legitimacy.[195] This assessment could change over time, potentially casting a long shadow over the Court's ability to maintain its legitimacy if wins are accumulated on one side of the political spectrum without sufficient offsetting.

---

189 *Id.* at 214–15.

190 *See* Katerina Linos & Kimberly Twist, *The Supreme Court, the Media, and Public Opinion: Comparing Experimental and Observational Methods*, 45 J. Legal Stud. 223, 247–49 (2016).

191 Logan Strother & Shana Kushner Gadarian, *Public Perceptions of the Supreme Court: How Policy Disagreement Affects Legitimacy*, 20 Forum 87, 104 (2022).

192 *Id.* at 98.

193 *Id.* at 97.

194 *Id.*; *see also* Gibson & Nelson, *supra* note 188, at 206–07 ("Individuals seem to keep a running tally of decisions, crediting the Court when it makes a pleasing decision and subtracting from the tally when the Court makes a disagreeable decision . . . [which] suggests that the Court's diffuse support could suffer once some accumulated threshold level of dissatisfaction is reached.").

195 Gibson & Nelson, *supra* note 188, at 211 ("[The Court's] legitimacy seems to flow from the view that discretion is being exercised in a principled, rather than strategic, way.").

2. *The Social Price of Preserving Legitimacy*

Even if the Court is able to shape the legitimacy consequences of its decisions, it is not necessarily able to avoid the social and legal consequences. In their paper *Establishment Clause Appeasement*, Professors Micah Schwartzman and Nelson Tebbe raise a series of concerns in that regard.[196] They argue that narrow decisions, while seeking to deescalate conflict[197] and limit the reach of the holding,[198] actually "work to incentivize an adversary's aggressiveness" and "confer legitimacy on [an] adversary's assertive position."[199] They further argue that voting with the majority, even to limit the reach of the holding, takes away dissenters and lends credence to the majority's view.[200] They note that appeasement can also move or widen the range of constitutional positions that are considered reasonable at any given moment[201] and ultimately can "do little to stop or slow the doctrinal inversions that we are witnessing under the Religion Clauses."[202]

The observed behavioral and legal outcomes of *Fulton* lend partial support to these concerns. *Fulton* appeared to have conferred legitimacy on foster care agencies that refuse service to same-sex couples for religious reasons, and to incentivize service refusal beyond litigants, and even beyond religious agencies. *Fulton* did succeed in reducing conflict in the short term, leading to the effective resolution of similar conflicts nationwide, but along the way it expanded the range of constitutional positions, as the spirit of *Fulton*—rather than its actual holding—controlled the resolution of cases with very different circumstances.[203]

Thus, our work points to a potential interplay between legitimacy consequences and social consequences. Strategically narrow decisions, designed to maintain legitimacy, are not without broader social and legal consequences. Whether these consequences are inadvertent or desirable from the perspective of the Court, and how they actually reflect on the Court's legitimacy, is hard to say. What is important, however, is that these consequences are now on the table. They enter the discussion of the pros and cons of the Roberts Court's specific brand of

196 Micah Schwartzman & Nelson Tebbe, *Establishment Clause Appeasement*, 2019 SUP. CT. REV. 271, 272 (2020) (introducing the concept of "appeasement," which they define as "a sustained strategy of offering unilateral concessions for the purpose of avoiding further conflict," as a tactic arguably deployed by liberal Justices in recent Establishment Clause cases).

197 *Id.* at 307.

198 *Id.* at 304.

199 *Id.* at 302.

200 *Id.*

201 *Id.* at 302–03.

202 *Id.* at 308.

203 *See supra* Part III.

incrementalism, which is minimalist in reasoning, but not in outcomes.[204] To the extent that narrow, incremental decision-making is favored by some members of the Court under the assumption that this approach will not have consequences exceeding the case at hand, our findings show that this assumption is false. Importantly, we have no direct access to several counterfactuals—a broad decision with a stinging dissent, a straightforward decision without any remedy, or, in between, a carefully tailored narrow decision that takes mitigating steps against overly broad interpretations. Whether such counterfactual decision-types could have had more or fewer social, legal, and legitimacy consequences is an issue for future articles. But our findings cast doubt on the notion that narrow decisions will have correspondingly narrow consequences.

Of course, none of this is to say that narrow decisions will *necessarily* be interpreted as broader than they are, in all circumstances. Rather, we suggest that our results are particularly relevant in several settings. First, they are relevant where a narrow decision is persuasively portrayed in relevant circles as a broad decision—a big, clear victory for one side. This is arguably easier when the holding is ambiguous, vague, or insincere (as both *Fulton* and *Masterpiece Cakeshop* were analyzed and criticized to be). Second, they are relevant where a narrow decision comes as part of a series of decisions whose outcomes build up a broad agenda. In such settings, the relevant stakeholders are likely to connect the dots and pay attention to the decision's outcome more than the seemingly narrow reasoning. In the present moment of free exercise doctrine, the Court appears to signal a broader agenda than the reasoning that each decision alone seemingly offers. On this point, it is worth quoting the address of Professor Barak-Corren to the *Fulton* court before the case was decided:

> [T]he Court has the power to shape public attitudes and public behavior, thereby producing either less or more bias and discrimination in society. It is true that after a decision is handed down by the Court, it takes on a life of its own, and much of its effects depend on how it is communicated by mass media. But the Court is not a helpless statist in this process. A clearer and less ambiguous decision—for example, one that sets a clear rule that is easy to communicate, understand, and follow—is less open to aggrandization, misstatements, or misinterpretations. The *Fulton* Court should opt for a clear and bright-line decision that provides specific and unambiguous behavioral

[204] For an opposite brand of judicial incrementalism that is minimal in outcomes, yet not in reasoning, see Rivka Weill, *The Strategic Common Law Court of Aharon Barak and its Aftermath: On Judicially-Led Constitutional Revolutions and Democratic Backsliding*, 14 L. & Ethics Hum. Rts. 227 (2020).

> instructions, including for future cases. This is particularly true if the court decides to grant an exemption in *Fulton*. In such case, the Court should assume, based on the *Masterpiece Cakeshop* effect, that its decision will likely encourage discrimination against sexual minorities. It is the Court's responsibility to minimize this effect to the extent possible. The Justices should not mislead themselves to think that evading the big questions or making a case-specific decision, as in *Masterpiece Cakeshop*, will avoid undesirable outcomes. The Justices should also not mislead themselves to think that their decision will only expand the freedom of a negligible minority of extremely objecting individuals. Rather, exempting religious objectors will likely have a broad impact, including on decision-makers who were willing to provide services before the decision, but will refuse to do so afterwards. These consequences are particularly concerning given the number of wedding conflict cases that have recently resolved in favor of vendors. This trend gives rise to the possibility that the *Masterpiece Cakeshop* effect will repeat and aggregate over time, the more such decisions are made and become known to the public.[205]

C. Incorporating Consequences into Constitutional Doctrine

After expanding the evidentiary basis of consequential constitutionalism and introducing legitimacy consequences to the discussion, we turn to show how the approach we developed in this paper can help improve the doctrinal application of constitutional law.

Given the context of this Article, we focus on free exercise doctrine and particularly on the compelling interest test, a central component of that doctrine. This test was first introduced to the religion context in *Sherbert v. Verner* as part of the three-prong strict scrutiny test.[206] Under this framework, a law that (1) "imposes any burden on the free exercise of . . . religion"[207] survives only if (2) it is justified by "some compelling state interest" and (3) there are no less restrictive means of fulfilling that interest.[208]

Sherbert controlled free exercise claims until *Employment Division v. Smith.*[209] *Smith* held that only a non-neutral law or a law that is not

---

[205] Barak-Corren, *A License to Discriminate?*, *supra* note 82, at 364 (footnotes omitted).

[206] 374 U.S. 398, 403, 406–09 (1963).

[207] *Id.* at 403. The "any burden" part was later modified into "undue" or "substantial" burden. *See, e.g.*, Wisconsin v. Yoder, 406 U.S. 205, 220 (1972) ("A regulation neutral on its face may, in its application, nonetheless offend the constitutional requirement for government neutrality if it unduly burdens the free exercise of religion.").

[208] *Sherbert*, 374 U.S. at 406–07.

[209] Emp. Div., Dep't of Hum. Res. v. Smith, 494 U.S. 872 (1990). Studies have shown, however, that while strict scrutiny was officially applied to free exercise claims, in reality,

generally applicable is subject to strict scrutiny.[210] *Smith* resulted in the application of rational basis review, a more deferential standard, to generally applicable laws that burden the free exercise of religion.[211]

Smith has been criticized since the day it was handed down[212] and through a decades-long evolution that included some back-and-forth between the Court and Congress, its holding has been gradually eroded. Today, federal legislation burdening free exercise is subject to strict scrutiny[213] and although *Smith* still controls state conduct[214] recently the Supreme Court has been shifting the relationship between its parts. While *Smith*'s two standards—rational basis review for generally applicable laws and strict scrutiny for non-neutral laws—are still ostensibly in use, the Court eroded the first group in *Fulton* by classifying Philadelphia's otherwise neutral antidiscrimination policy as not generally applicable due to a never-used and arguably irrelevant contractual mechanism for granting individual exemptions.[215] In addition

courts were often hesitant to grant religious exemptions. *See, e.g.*, Ira C. Lupu, *Where Rights Begin: The Problem of Burdens on the Free Exercise of Religion*, 102 Harv. L. Rev. 933, 947–48 (1989) (describing why free exercise claims are "often deeply troubling").

210 *Smith*, 494 U.S at 885. For a taxonomy of the categories of cases that fall under strict scrutiny after *Smith*, see W. Cole Durham & Robert Smith, 1 Religious Organizations and the Law § 3:7 (2d ed. 2023) (outlining five such categories: "(1) cases involving religious belief, (2) cases dealing with church autonomy," (3) cases examined under the new doctrine of hybrid rights "(where free exercise and another constitutional right are both being violated), (4) cases in which individual assessments are the focus (where the applicability of a law to an individual is being adjudged), and (5) cases in which the laws fall short of neutrality or general applicability by targeting religion").

211 An empirical examination of federal decisions published between 1990 and 2003 showed that laws that were challenged on religious liberty grounds had the highest survival rate of any category of challenge in which strict scrutiny was sought (fifty-nine percent as compared with a thirty percent average survival rate). The research distinguished between generally applicable laws, that survived seventy-four percent of the time even if they did not provide religious exemptions, and laws that targeted religion in a non-neutral manner, which had zero chance of survival. Adam Winkler, *Fatal in Theory and Strict in Fact: An Empirical Analysis of Strict Scrutiny in the Federal Courts*, 59 Vand. L. Rev. 793, 859–61 (2006).

212 *See generally* James D. Gordon, *Free Exercise on the Mountaintop*, 79 Calif. L. Rev. 91 (1991); Michael W. McConnell, *Free Exercise Revisionism and the* Smith *Decision*, 57 U. Chi. L. Rev. 1109 (1990); Douglas Laycock, *The Remnants of Free Exercise*, 1990 Sup. Ct. Rev. 1 (1990).

213 Congress responded to *Smith* by codifying strict scrutiny through the Religious Freedom Restoration Act (RFRA). Pub. L. No. 103-141, 107 Stat. 1488 (1993) (codified at 42 U.S.C. §§ 2000bb to 2000bb-4). Congress subsequently enacted the Religious Land Use and Institutionalized Persons Act (RLUIPA), which codified strict scrutiny review for burdens placed on religious exercise in cases involving prisoners and land use. Pub. L. No. 106-274, 114 Stat. 803 (2000) (codified at 42 U.S.C. §§ 2000cc to 2000cc-5).

214 RFRA initially applied to both the federal government and the states; however, in *City of Boerne v. Flores*, 521 U.S. 507, 533–36 (1997), the Court limited RFRA's application to the federal government.

215 *See* Fulton v. City of Philadelphia, 141 S. Ct. 1868, 1876–79 (2021) (arguing that "[a] law is not generally applicable if it . . . provid[es] a mechanism for individualized exemption"); *id.*

to narrowly interpreting the category of generally applicable laws, five Justices signaled willingness to overturn *Smith* altogether and restore strict scrutiny to its previous prominence.[216] Whatever fate lies ahead for *Smith*, strict scrutiny already plays and will continue to play a central role in the future of free exercise cases.

1. The Problematic Application of the Compelling State Interest Test

Courts and scholars have consistently struggled to define compelling state interests. Justice Blackmun observed more than forty years ago that "I have never been able fully to appreciate just what a 'compelling state interest' is,"[217] a concern Justice Stevens echoed.[218] Professor Richard Fallon noted that "the generative cases that launched the strict scrutiny test said nothing about how courts should identify compelling interests. In subsequent cases, too, the Supreme Court has frequently adopted an astonishingly casual approach in labeling asserted governmental interests as either compelling or not compelling."[219] Professors Douglas Laycock and Oliver Thomas complained that "[t]he compelling interest test has fallen into disarray, especially in the lower courts."[220]

The Court's lack of clarity about what makes a state interest compelling was highlighted in *Fulton*. The Court dismissed Philadelphia's interests in maximizing the number of foster parents and ensuring equal treatment as overly general and unfounded.[221] And while Justice Alito opined in his concurrence that the government must "show with . . . particularity how [that interest] . . . would be adversely affected by

---

at 1929 (Gorsuch, J. concurring) (questioning whether the majority's focus on the contract's provision for individualized exemptions is "just stretching to claim some cover for its novel [legal] arguments").

216 It is unclear if *Smith* were to be overturned, what exactly it would be replaced with. In *Fulton*, Justices Alito, Thomas and Gorsuch called for overturning *Smith* and restoring "the standard that *Smith* replaced: A law that imposes a substantial burden on religious exercise can be sustained only if it is narrowly tailored to serve a compelling government interest." *Id.* at 1924 (Alito, J., concurring). Justices Barrett and Kavanaugh were critical of *Smith* but also hesitant of the notion that it should be replaced with a categorical strict scrutiny regime. *Id.* at 1883 (Barrett, J., concurring).

217 Ill. Elections Bd. v. Socialist Workers Party, 440 U.S. 173, 188 (1979) (Blackmun, J., concurring).

218 Eu v. S.F. Cnty. Democratic Cent. Comm., 489 U.S. 214, 233 (1989) (Stevens, J., concurring).

219 Richard H. Fallon, Jr., The Nature of Constitutional Rights: The Invention and Logic of Strict Judicial Scrutiny 54 (2019).

220 Douglas Laycock & Oliver S. Thomas, *Interpreting the Religious Freedom Restoration Act*, 73 Tex. L. Rev. 209, 222 (1994).

221 *Fulton*, 141 S. Ct. at 1881–82 (2021).

granting an exemption,"[222] the Court offered the parties no guidelines for how to substantiate their interests at the right level of precision. The Court's rejection of each of these interests was extremely brief—one or two sentences each—explaining the City "fail[ed] to show" and "offer[ed] no compelling reason" why granting CSS an exception would put those goals at risk.[223] The Court did not explain what kind of evidence the City could have brought to make its argument concrete and compelling.

Although the Court has been unclear about what exactly makes a state interest compelling, the case law makes it clear that the state needs to provide evidence of its compelling interest. Abstract or general claims will generally not suffice.[224] Even where a state interest is

---

[222] *Id.* at 1890 (Alito, J., concurring) (quoting Wisconsin v. Yoder, 406 U.S. 205, 236 (1972)); *see also* Gonzales v. O Centro Espirita Beneficente Uniao do Vegetal, 546 U.S. 418, 431 (2006) (noting that, when applying the compelling interest test, the Court should look "beyond broadly formulated interests justifying the general applicability of government mandates and scrutinize[] the asserted harm of granting specific exemptions to particular religious claimants"); Holt v. Hobbs, 574 U.S. 352, 363 (2015) (explaining that the Court has to look at the interest in enforcing the policy in the specific context).

[223] *Fulton*, 141 S. Ct. at 1881–82. The rejection of the equality interest hinged on the Court's interpretation of the Philadelphia contract as allowing discretionary exemptions. *Id.* at 1878. However, as Justice Gorsuch noted, the contractual provision was "irrelevant" to the dispute and was not part of the arguments. *Id.* at 1928–29 (Gorsuch, J., concurring). Nor did the City ever grant anyone an exemption from this provision. *See* Ira C. Lupu & Robert W. Tuttle, *Two Surprises in* Fulton v. City of Philadelphia*—A Unanimous Outcome and the Enduring Quality of Free Exercise Principles*, Am. Const. Soc'y (June 18, 2021), https://www.acslaw.org/expertforum/two-surprises-in-fulton-v-city-of-philadelphia-a-unanimous-outcome-and-the-enduring-quality-of-free-exercise-principles [https://perma.cc/6TNH-DCYN]. The Court also rejected a third interest in minimizing liability for potential lawsuits in one sentence. *Fulton*, 141 S. Ct. at 1881–82 ("[M]inimizing liability . . . [is an] important goal[], but the City fails to show that granting CSS an exception will put those goals at risk.").

[224] For example, in *Sherbert v. Verner*, the state claimed that it had a compelling interest in denying the claimant unemployment benefits based on the possibility that unscrupulous claimants could feign religious objections to working on Saturday, which would dilute the unemployment fund, and hinder employees from scheduling necessary work on Saturday. 374 U.S. 398, 406–07 (1963). The Court did not acknowledge this asserted interest as compelling because the state offered no evidence or proof and did not even raise the claim before reaching the Supreme Court. *Id.* at 407–08. This reasoning fits within the larger context of equal protection jurisprudence, which makes clear that there is no generalized compelling interest in remediating discrimination absent evidence of specific discriminatory acts. *See, e.g.*, Wygant v. Jackson Bd. of Educ., 476 U.S. 267, 276–77 (1986) ("Societal discrimination, without more, is too amorphous a basis for imposing a racially classified remedy . . . [the actor] must have sufficient evidence to justify the conclusion that there has been prior discrimination."); Bush v. Vera, 517 U.S. 952, 982 (1996) (explaining that a State's interest in remedying discrimination is compelling where (1) the discrimination that the State is specific and identified and (2) the State has "a strong basis in evidence" to conclude that "remedial action was necessary"). The theoretical underpinnings of this idea are attributed in this context to J. Morris Clark. *See* J. Morris Clark, *Guidelines for the Free Exercise Clause*, 83 Harv. L. Rev. 327, 330–31 (1969) (Because the "purpose of almost any law can be traced back to . . . fundamental concerns of government: public health and safety, public peace and order,

compelling in general, the government bears the burden to prove that it is compelling as applied to the specific circumstances of the case and exemption sought.[225]

While the Justices acknowledge the importance of empirical evidence, they do not agree on how to interpret and apply it. For example, in *Wisconsin v. Yoder*, a seminal free exercise case, Chief Justice Burger writing for the majority, Justice White in his concurrence, and Justice Douglas in his dissent, all referenced different social science studies.[226] *Yoder* dealt with defendants from the Amish community who were found guilty of violating a compulsory education law by refusing to send their children to school beyond the eighth grade.[227] The majority relied on a testimony offered by a professor of education who detailed the benefits of the Amish vocational education; the Court accepted his description of the Amish education as "ideal."[228] Justice White referred to another testimony regarding the widespread attrition from the Amish community.[229] Justice Douglas denied that there was enough evidence on the issue, arguing that the views of Amish children in Wisconsin should be canvassed and referencing other empirical social science research to emphasize children's right to make decisions concerning their education.[230] While the Justices recognized the importance of evidence, they neither agreed about how nor had the tools to decide what makes evidence relevant and persuasive.

More recently, in a free speech case, *Brown v. Entertainment Merchants*, the Supreme Court analyzed whether a law that imposed restrictions and labeling requirements on violent video games sold to minors served a compelling state interest.[231] Justice Scalia, writing for

---

defense, revenue," courts should compare individual interests to "the more particular values served by the chosen regulation" rather than "such highly generalized secular values").

225 In their survey of 264 federal cases at all levels between 1990 and 2015 in which religious exercise cases were adjudicated under a strict scrutiny framework, Wolanek and Liu found that different compelling interests had different likelihoods of success. Caleb C. Wolanek & Heidi H. Liu, *Applying Strict Scrutiny: An Empirical Analysis of Free Exercise Cases*, 78 MONT. L. REV. 275, 298–300 (2017). The authors show that rarely are government interests rejected categorically as not compelling. *Id.* at 304–07. In most cases, courts deem the interest not compelling *as applied* to the specific claimant, implying that a more directly demonstrated governmental interest could have led the case to a different path. *Id.* at 305. This finding indicates that given more directly relevant evidence, courts might rule differently on the compelling interest question.

226 *See* Wisconsin v. Yoder, 406 U.S. 205, 223–24 (1972); *id.* at 240 n.3 (White, J., concurring); *id.* at 245 n.3 (Douglas, J., dissenting in part).

227 *Id.* at 207–09 (majority opinion).

228 *Id.* at 223–24.

229 *Id.* at 240 n.3 (White, J., concurring).

230 *Id.* at 246 & nn.3–4 (Douglas, J., dissenting in part).

231 Brown v. Ent. Merch. Assoc., 564 U.S. 786, 799 (2011).

the Court, argued that the law failed the compelling interest test.[232] While he acknowledged there was a showing that violent video games led to violent behavior, he explained that the state could not "show a direct causal link between violent video games and harm to minors."[233] Furthermore, Scalia argued that "the government [did] not have a compelling interest in each marginal percentage point by which its goals are advanced,"[234] suggesting that the effects in question need to be substantial and not marginal in order to be considered compelling.

Justice Breyer's dissent featured two lengthy appendices that listed peer-reviewed academic articles on the psychological harms resulting from playing violent video games.[235] He compiled the lists with the assistance of the Supreme Court Library,[236] and explained that "[s]ocial scientists, for example, have found *causal* evidence that playing these games results in harm."[237] Although he admitted to lacking the expertise to evaluate conflicting evidence, he concluded, "[u]nlike the majority, I would find sufficient grounds in these studies and expert opinions for this Court to defer to an elected legislature's conclusion that the video games in question are particularly likely to harm children."[238]

The disagreement between Justices Breyer and Scalia highlights the tension between the Court's *need* for evidence to apply the doctrine and its lack of a *method* to evaluate the merits of the evidence.[239] The Court seems to treat social science data instrumentally, often choosing to rely on evidence or disregard it rather than attempting to adjudicate contradictory evidence and evaluate the weight and persuasiveness

---

232 *Id.* at 803–04.

233 *Id.* at 799.

234 *Id.* at 803 n.9.

235 *Id.* at 858–72 (Breyer, J., dissenting). For an analysis of Justice Breyer's dissent, see Martha Minow, *The Big Picture: Justice Breyer's Dissent in* Brown v. Entertainment Merchants Association, 128 Harv. L. Rev. 469 (2014).

236 *Ent. Merchs. Assoc.*, 564 U.S. at 858 (Breyer, J., dissenting).

237 *Id.* at 851.

238 *Id.* at 855.

239 Confronted with the same evidence, Justices often do not agree on what it means. *Compare* Burwell v. Hobby Lobby Stores, Inc., 573 U.S. 682, 726 (2014) (Alito, J.) (criticizing the government's assertion that a "contraceptive mandate serves a variety of important interests" as being "couched in very broad terms"), *with id.* at 761 (Ginsburg, J., dissenting) ("[T]he Government has shown that the contraceptive coverage for which the [Affordable Care Act] provides furthers compelling interests in public health and women's well being. Those interests are concrete, specific, and demonstrated by a wealth of empirical evidence."). *See also* Emp. Div., Dep't of Hum. Res. v. Smith, 494 U.S. 872, 911–15 (1990) (Blackmun, J., dissenting) (noting the lack of evidence offered by the majority to prove the inherent danger of peyote and offering empirical evidence of the positive impacts of the use of peyote in a religious context and the lack of illegal traffic in peyote).

of studies that are directly relevant to asserted compelling interests.[240] The result is an unprincipled decision making that simultaneously requires evidence and often ignores it.

2. *Concretizing Compelling Interests Empirically*

Although this Article cannot deal with the diverse universe of cases, problems, and circumstances that require the assessment of compelling interests and balancing more generally, our study of foster care agencies and their treatment of same-sex couples provides two methodological starting points to concretize the compelling interest test in religion-equality conflicts. Furthermore, our results are particularly useful given the Court's recent interest in examining state practice as part of the compelling interest test.[241] We provide accumulative knowledge from all states on both their policies and their outcomes—exactly the kind of evidence in which the Court is interested.[242]

First, we provide empirical evidence on the compelling interest question in *Fulton*. Recall that the City alleged that it had a compelling interest in maximizing the number of foster parents and ensuring the equal treatment of LGBTQ+ families.[243] Both of these interests are harmed if parents are turned down because of their sexual orientation. To show the importance of these goals, the City could have provided evidence on the differential disparity in denials of service to

240 For example, in *Yoder*, the Justices referred to different studies and took away different messages from the experts who testified. *See supra* notes 226–30 and accompanying text. Likewise, in *South Bay United Pentecostal Church v. Newsom*, 141 S. Ct. 716 (2021), the Justices could not agree either on the science or how the science fit with the First Amendment. Recall that in *South Bay United Pentecostal Church*, the Court partially lifted California's restrictions on religious services. *See id.* at 716. Chief Justice Roberts's concurrence explained that he saw "no basis" for "overriding" the state's conclusion that singing indoors poses an increased risk of transmitting COVID-19 but dismissed California's conclusion that it was unsafe to worship in the "cavernous cathedral" on the grounds that it reflected neither "expertise" nor "discretion." *Id.* at 717 (Roberts, C.J., concurring). Justice Gorsuch's concurrence accused California of worrying about the health risks of worship while still allowing its citizens to sit on buses or get a manicure. *Id.* at 718–19 (Gorsuch, J., concurring). But Justice Kagan's dissent accused the majority of "displac[ing] the judgments of experts about how to respond to a raging pandemic" and cited the testimony of a public health expert on the risks of singing and indoor gatherings. *Id.* at 720–22 (Kagan, J., dissenting). Indeed, the expert explained why some activities, like shopping, are less risky than gatherings at churches. *Id.* at 721–22.

241 *See, e.g.*, Ramirez v. Collier, 142 S. Ct. 1264, 1286–88 (2022) (examining Texas's practice of banning religious advisors from being able to physically touch inmate during the execution process).

242 *Id.* at 1287 (discussing how, to tell what a "compelling" state interest is, the Court "sometimes . . . looks to a State's policy-based or commonsense arguments" and "[o]ften also examines history and contemporary state practice").

243 Fulton v. City of Philadelphia, 141 S. Ct. 1868, 1881 (2021).

same-sex couples in jurisdictions with and without religious exemptions. Uncovering this evidence is one of the contributions of our work.[244]

The second element that *Fulton* allows us to study is the behavioral concern that exemptions could increase the demand for exemptions or encourage undesirable behavior. In brief, the Court often asks whether granting an exemption to one claimant could open the floodgate for other claimants and erode the rule of law.[245] The Court has at times accepted and at times rejected concerns that an exemption will erode the rule, without demarcating a clear line between persuasive and unpersuasive versions of the argument and without offering a methodology to mark this line.[246] This is a question we can study empirically. If demand for exemptions is fixed, no one who previously did not want an exemption would claim one in response to its new availability. Conversely, if the demand is elastic, religious exemptions can encourage behavior change. If the availability of exemptions encourages discriminatory behavior, the government interest in denying exemptions is stronger. Our results suggest that this concern should be weighed more seriously in the future because demand for exemptions appears to be elastic and responsive to the legal environment.[247]

Our method can be extended and replicated by any state or city that is faced with a need to evaluate its policy as applied to specific religious claimants. States can launch their own studies, search for existing patterns of discrimination, and evaluate their respective interests in enforcing their laws uniformly using relevant data.

Notably, we find substantial variation between states. Some states contract with dozens of agencies, others contract with only a handful. In

---

244 *See supra* Part II.

245 *See, e.g.*, Braunfeld v. Brown, 366 U.S. 599, 609 (1961) (expressing concern that granting religious exemptions to Pennsylvania law requiring all businesses to close on Sundays could lead to a proliferation of similar claims).

246 The Court has primarily accepted the argument in cases involving large administrative systems. *See, e.g.*, Hernandez v. Comm'r, 490 U.S. 680, 699, 703 (1989) (taxation); United States v. Lee, 455 U.S. 252, 261 (1982) (social security). In the past, the Court has also insisted that no exemptions can be provided from a nondiscrimination policy, seemingly accepting the concern that exemptions will erode the rule. *See* Bob Jones Univ. v. United States, 461 U.S. 574 (1983) (holding that nonprofit private schools that enforce racially discriminatory religious doctrine do not qualify as tax-exempt organizations). However, the Court rejected the argument in the context of unemployment benefits in *Sherbert v. Verner*, 374 U.S. 398, 412–13 (1963), and in other decisions. *See* Gonzales v. O Centro Esperita Beneficente Uniao do Vegetal, 546 U.S. 418, 436 (2006) ("The Government's argument echoes the classic rejoinder of bureaucrats throughout history: If I make an exception for you, I'll have to make one for everybody, so no exceptions. But RFRA operates by mandating consideration, under the compelling interest test, of exceptions to rule[s] of general applicability."); Burwell v. Hobby Lobby Stores, Inc., 573 U.S. 682, 717–18 (2014) (rejecting the government's argument that federal courts will be unable to "weed out insincere claims").

247 *See supra* Part II.

some states, religious agencies dominate the market, whereas in other states they are a minority. In many states we find that discrimination is prevalent, whereas in a few states it is rare. How these differences should be factored into the analysis is a complicated matter. In general, states' interest in ensuring equality appears stronger when discrimination is substantial and/or when the objecting agencies (whether one or many) are responsible for a large share of the foster care services market. These factors can calibrate the analysis empirically within each state. We acknowledge that there are many additional factors that our study did not uncover and can be added to the discussion,[248] but our study illustrates the potential of documenting and assessing this variation.

3. Testing the Means to Ensure Narrow Tailoring

To know whether a universal enforcement of antidiscrimination laws is the least restrictive means to ensure equal treatment and access to public services, courts need to know whether religious exemptions detract from this compelling goal. The results of the *Masterpiece Cakeshop* and *Fulton* field experiments establish that both decisions detracted from this goal in most jurisdictions by substantially expanding discrimination against same-sex couples.[249] In addition, although *Fulton* (and *Masterpiece Cakeshop*) did not have negative effects on jurisdictions that already enacted religious exemption laws to their antidiscrimination rules, pre-*Fulton* results show that these jurisdictions were associated with the highest rates of discrimination.[250] These combined results indicate that religious exemption laws, in their current form, are not means that achieve the compelling interest in ensuring equality. Since governments' duty under strict scrutiny to choose the least restrictive means pertains to means that are effective in achieving the compelling interest, allowing religious exemptions, at least as currently designed, does not appear to be a satisfying solution.

An important implication of the results of the *Masterpiece Cakeshop* and present *Fulton* experiments is that efforts to find means to reduce religion-equality conflicts should rely on (reliable and robust) empirical evidence regarding the likely consequences of the proposal. Only by measurement can policymakers ensure that their means are narrowly tailored to the compelling interest they seek to advance. It is possible that a different combination of legal means than those applied to *Masterpiece Cakeshop* and *Fulton* would generate different

248 *See supra* Section II.B.I.
249 *See supra* Sections IV.A.I–II.
250 *See supra* Section II.B.I.

behavioral outcomes and be better at balancing religious freedom and equality. Pre-measurement by experimentation provides the safest way to conduct this measurement and tailor the means by trial and error.

To that effect, Professor Barak-Corren has proposed to pre-test the likely effects of proposed policies before formal implementation.[251] She has suggested that a legislature could collect a representative sample of the state population, randomly expose different groups to alternative bills, and examine whether exposure to one bill (compared to the others, or no bill) generates more or less antigay bias in the population, or produces a more or less accurate understanding of appropriate behavior.[252] Lawmakers could either devise their own decisionmaking dilemmas to probe citizens' understanding of a proposed law, or they could rely on one of the many measures established in psychological research to capture bias and social norm perceptions.[253]

Such experimentation can provide an answer to a lingering and consequential question: are there means that lie between uniform application of the law and broad religious refusal laws? Can such means be tailored to achieve equal treatment while preserving religious freedom to the maximal attainable degree?[254] This investigative avenue

---

[251] In general, the field of evidence-based lawmaking has been developing theoretically and practically in recent years. *See, e.g.*, Ittai Bar-Siman-Tov, *The Dual Meaning of Evidence-based Judicial Review of Legislation*, 4 Theory & Prac. Legis. 107, 109 (2016) (noting that "'evidence-based judicial review of legislation' is a still undeveloped and rarely-used term" and offering two new conceptualizations of the topic).

[252] *See* Barak-Corren, *A License to Discriminate?*, *supra* note 82, at 360–61.

[253] The literature on capturing bias and social norms perceptions is huge. *See, e.g.*, Ofosu, Chambers, Chen & Hehman, *supra* note 173 (using approximately one million responses to an implicit association test over a twelve-year window to examine antigay bias before and after the legalization of same-sex marriage); Tankard & Paluck, *supra* note 171 (finding an increase in perceived social norms after the legalization of gay marriage but no change in personal attitudes).

[254] Professor Barak-Corren explored a set of such measures in the context of healthcare coverage. *See* Netta Barak-Corren, *The War Within Religion: Towards a More Nuanced Resolution of Religion/Equality Conflicts*, 71 Am. J. Compar. L. (forthcoming 2024) (manuscript at 32–33), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3183733. Faced with religious objections to providing health benefits to same-sex partners, San Francisco and Michigan found a creative solution: instead of specifying the beneficiaries as same-sex partners, employees were allowed to designate any member of their household to receive health benefits, regardless of the exact nature of the relationship. This solution was successful in removing religious objections and securing wide compliance. *Id.* (manuscript at 36–37) (discussing compliance). For an example of an attempt to strike a similar balance in foster care, see 2023 Utah Laws Ch. 466, a recent law that mandates all foster care agencies in Utah to join a consortium wherein agencies have the right to refer prospective parents to other agencies in the consortium, but not to deny service to a couple. The Utah solution is not perfect—it would have been preferable if couples would never have to face any rejection, including by referral. A better option would have been to channel all couple inquiries to one address in the consortium, which would provide couples with a list of able and willing agencies after eliminating all that would or could not provide the service, for religious or

is open not only to religion-equality conflicts, but, with the appropriate modifications, can be applied to a wide range of constitutional issues where strict scrutiny may apply to government decision making.

D. Limitations of Consequential Constitutionalism

Empirical explorations are always limited in what the methods can and cannot reveal; as Richard Posner noted, not all data of relevance and interest *can* be unearthed and some uncertainty is likely to endure.[255] Our work is no different. First, we examine only the first stage of the licensing process. Further stages of the process involve additional stages of evaluation of prospective foster families; disparate treatment can arise in many stages down the line, including in the interview phase, in home studies, and in the final placement of children in families. We do not test these additional stages and hence can offer no evidence on the total or final consequences of discrimination. Naturally, disparities that arise in the first contact phase reduce the number of options available for couples further down the line.

Second, our experimental design is focused on the short-term effects of *Fulton* and cannot shed light on possible long-term effects of the decision, including whether the short-term effects persist. Any effect could either attenuate or strengthen in the long term due to intervening events, such as a social or legal backlash and legal events inspired by the decision (whether acts of legislation, new litigation, new decisions, and so on). We partially address this limitation by investigating the outcomes of all similar cases that were pending at the time *Fulton* was decided. While this analysis cannot reveal the full nationwide scope of the long-term effects of the decision, it provides important information on how *Fulton* shaped the legal landscape onwards.

Third, our work does not unearth all of the considerations that factor into the balancing of conflicting individual rights, as well as between individual rights and governmental interests. Some outcomes and harms resist measurement, others cannot be aggregated easily, and other consequences and preferences might yield inconclusive results. The replicability crisis in the social sciences[256] also makes scientists acutely

---

other reasons. She thanks Robin Fretwell Wilson for discussing this example with her, and for working relentlessly on finding solutions for these tough problems.

255 *See* Posner, *supra* note 1, at 12.

256 *See* Krin Irvine, David A. Hoffman & Tess Wilkinson-Ryan, *Law and Psychology Grows Up, Goes Online and Replicates*, 15 J. Empirical Legal Stud. 320, 322–23 (2018) (recommending that researchers are transparent in "piloting, public data storage, multiple scenario testing, and diverse platform subject series," to "shore up thefoundations of law and psychology"); Edith Beerdsen, *Litigation Science After the Knowledge Crisis*, 106 Cornell L. Rev. 529, 584 (2021) (noting that the replicability crisis has "far-reaching implications" in

aware of the difficulty of generalizing from any single empirical finding, and taught us the importance of techniques such as pre-registration of studies and data transparency to partially address this issue.[257]

Factoring real, measurable consequences into constitutional doctrine poses a real challenge. In addition to the limitations and extra care required by the empirical methods themselves, consequential constitutionalism might result in outcomes that advocates and stakeholders may not like. These could be teaching moments for the better or worse. Sometimes we will need to accept that our moral intuitions might have weaker empirical grounding than expected, and on other occasions such findings will prompt us to carefully consider whether (or when) consequences really are what we should care about when evaluating constitutional rules.

We do not believe empirical examinations exhaust constitutional debates. Quite the contrary. As Professor Barak-Corren argued multiple times, the value of empirical examinations of constitutional law lies in its adding nuance to debates that are often viewed as intractable culture war battles or zero-sum games.[258] Empirical evidence can generate new understandings of the potential outcomes of different legal paths under consideration and can help to bring the parties closer together by establishing a common place to stand. As such, empirically-grounded consequential constitutionalism could help to ground courts better in the real world, lend credibility (when done properly) to controversial decisions, help Justices battle their natural inclination to decide cases based on their values and dispositions, and lead the Court away from decisions that would have dangerously unsettling outcomes for society. We have shown some of these advantages in our analysis of the consequentialist approach to the application and development of free exercise doctrine. Understanding the potential and limitations of empirical evidence is a task broader than the current Article, and it should be taken seriously. Empirical constitutional law can generate doctrinal and institutional insights about deference, procedures, and the methods of application of general doctrines to concrete cases, all in ways that can improve constitutional doctrine going forward.

---

litigation "for the role of a judge, the balance between the parties, and the standards by which we judge admissibility").

257 As noted above, the present study was pre-registered in aspredicted.org; its anonymized data was made publicly available; the results were triangulated with findings from a simultaneous longitudinal survey experiment; and the results replicated and expanded previous studies, including the *Masterpiece Cakeshop* field experiment. Overall, the current study adds robustness to previous findings regarding the social and behavioral effects of Supreme Court decisions in general, and particularly in the domain of religion-equality conflicts.

258 *See, e.g.*, Barak-Corren, *supra* note 254; Barak-Corren, *supra* note 97, at 59.

Conclusion

In this Article, we lay ground for empirically studying constitutional consequences by focusing on *Fulton v. City of Philadelphia*. We investigate the consequences of *Fulton* in three pertinent audiences: foster care agencies, the general public, and parties to similar conflicts. We find that *Fulton*, despite its appearance as a fact-specific, narrow opinion, increased discrimination against same-sex couples both in places where the decision was expected to modify the law and in places where it should have had no impact. *Fulton* also increased support in the general public for religiously-motivated refusal of service to same-sex couples. Finally, *Fulton* yielded a 100% win record for the cause of religious refusals in lower courts across the nation. These findings suggest that the Supreme Court's holdings can have a powerful expressive function and shape the behavior of key stakeholders. Our results also allow us to elucidate longstanding doctrinal debates: whether the government has a compelling state interest in burdening religious exercise in favor of preventing sexual orientation discrimination, and what it takes to answer whether the means are narrowly tailored to that interest.

At a time when American constitutional law searches for consistency, neutrality, and legitimacy, we propose a methodology that takes constitutional doctrines seriously and elucidates the real-world effects of strategic judicial decisionmaking, while recognizing the limitations of making constitutional doctrine turn on concrete consequences. The answers that empirically-grounded consequential constitutionalism provide are not uniform or simple, the method exposes the limitations of legal assumptions and human reasoning, and the overall legal outcomes are likely to be more nuanced, limited, and modest. But all of that is good news, we think, as it can produce a better—more rigorously grounded—constitutional law.

APPENDIX

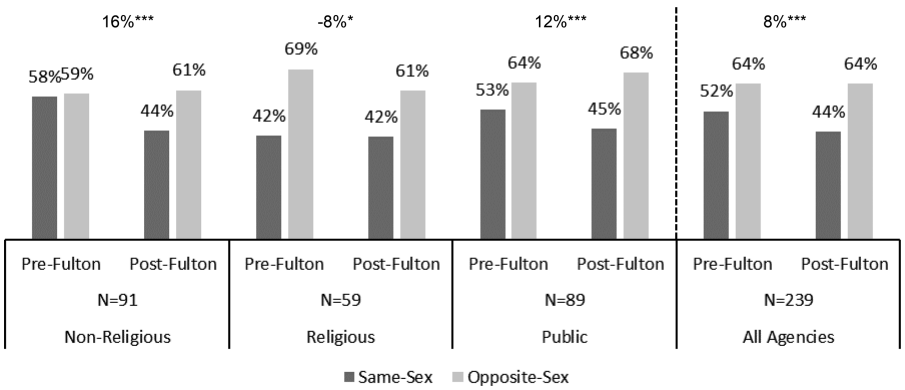
A. Additional Figures

TABLE A1. THE SHARE OF RELIGIOUS AGENCIES ACROSS LEGAL REGIMES

	Religious Exemption		No Religious Exemption	
	AD Rules, Michigan-like	**No AD Rules, Texas-like**	**AD Rules, Minnesota-like**	**No AD Rules, Ohio-like**
Number of agencies	67	239	743	185
Number of religious agencies	33	59	99	39
Share of religious agencies (%)	49%	25%	13%	21%
Mean percentage disparity between couples (%)	12%	14%	4%	13%

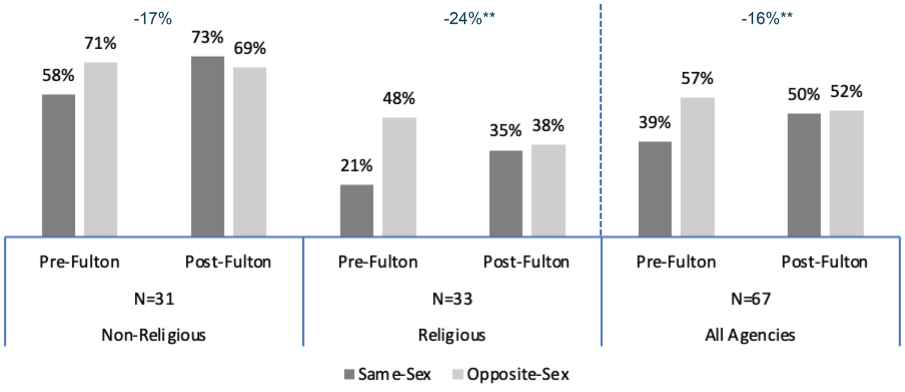
Note: AD = antidiscrimination

FIGURE A1. POSITIVE RESPONSE RATES TO SAME-SEX AND OPPOSITE-SEX COUPLES BY TYPE OF FOSTER CARE AGENCY IN TEXAS-LIKE JURISDICTIONS, BEFORE AND AFTER *FULTON*



* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The floating percentages represent the difference-in-differences from pre- to post-*Fulton*.

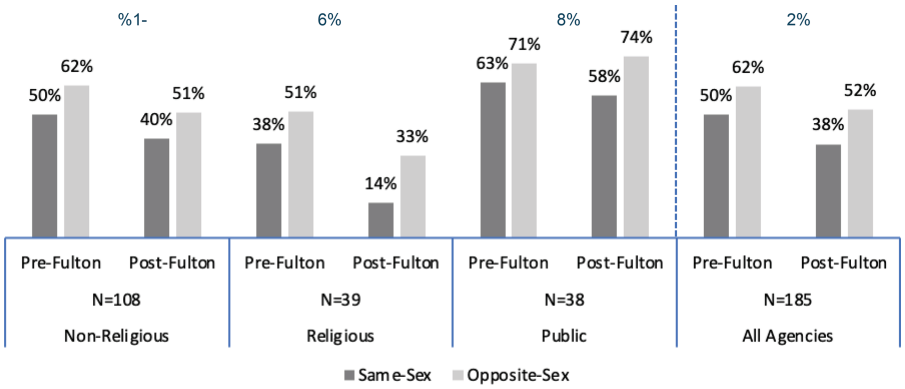
FIGURE A2. POSITIVE RESPONSE RATES TO SAME-SEX AND OPPOSITE-SEX COUPLES BY TYPE OF FOSTER CARE AGENCY IN MICHIGAN-LIKE JURISDICTIONS, BEFORE AND AFTER *FULTON*



* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: The response rate for public agencies could not have been calculated as there were only three such agencies in this cohort. The floating percentages represent the difference-in-differences from pre- to post-*Fulton*.

FIGURE A3. POSITIVE RESPONSE RATES TO SAME-SEX AND OPPOSITE-SEX COUPLES BY TYPE OF FOSTER CARE AGENCY IN OHIO-LIKE JURISDICTIONS BEFORE AND AFTER *FULTON*



* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The floating percentages represent the difference-in-differences from pre- to post-*Fulton*.

TABLE A2: STATES' SHARE OF RELIGIOUS AGENCIES AND AVERAGE RATES OF DISCRIMINATION

State	AD rules	RE rules	Number of agencies in final sample (excluding control)	Percent religious	Rate of same-sex discrimination
AL	No	Yes	83	13%	16%
AK	No	No	1	0%	0%
AZ	Yes	No	22	23%	2%
AR	No	No	6	33%	30%
CA	Yes	No	101	11%	5%
CO	Yes	No	51	8%	-3%
CT	Yes	No	15	20%	-14%
DE	Yes	No	3	33%	-17%
DC	Yes	No	4	0%	-26%
FL	Yes	No	15	27%	-7%
GA	No	No	65	32%	8%
HI	Yes	No	1	0%	0%
ID	Yes	No	3	33%	5%
IL	Yes	No	29	28%	18%
IN	Yes	No	21	5%	-4%
IA	Yes	No	10	50%	12%
KS	No	Yes	13	0%	6%
KY	Yes	No	32	22%	20%
LA	No	No	9	11%	41%
ME	Yes	No	2	100%	33%
MD	Yes	No	6	67%	18%
MA	Yes	No	21	10%	-1%
MI	Yes	Yes	40	53%	15%
MN	Yes	No	73	1%	12%
MS	No	Yes	3	33%	-17%
MO	Yes	No	21	19%	-15%
MT	Yes	No	5	20%	-30%
NE	Yes	No	6	50%	-10%
NV	Yes	No	3	33%	-15%
NH	Yes	No	5	0%	5%
NJ	Yes	No	9	22%	26%

State	AD rules	RE rules	Number of agencies in final sample (excluding control)	Percent religious	Rate of same-sex discrimination
NM	Yes	No	8	13%	-8%
NY	Yes	No	42	12%	10%
NC	No	No	19	37%	20%
ND	No	Yes	1	0%	-100%
OH	No	No	66	11%	9%
OK	No	Yes	11	36%	19%
OR	Yes	No	1	0%	-50%
PA	Yes	No	94	15%	8%
RI	Yes	No	5	40%	18%
SC	No	Yes	5	80%	-9%
SD	Yes	Yes	7	57%	-6%
TN	Yes	Yes	20	40%	17%
TX	No	Yes	65	35%	18%
UT	No	No	23	4%	13%
VT	Yes	No	2	50%	25%
VA	No	Yes	57	28%	9%
WA	Yes	No	20	10%	5%
WV	Yes	No	14	14%	4%
WI	Yes	No	76	3%	1%
WY	No	No	20	0%	6%

Note: AD = antidiscrimination; RE = religious exemption

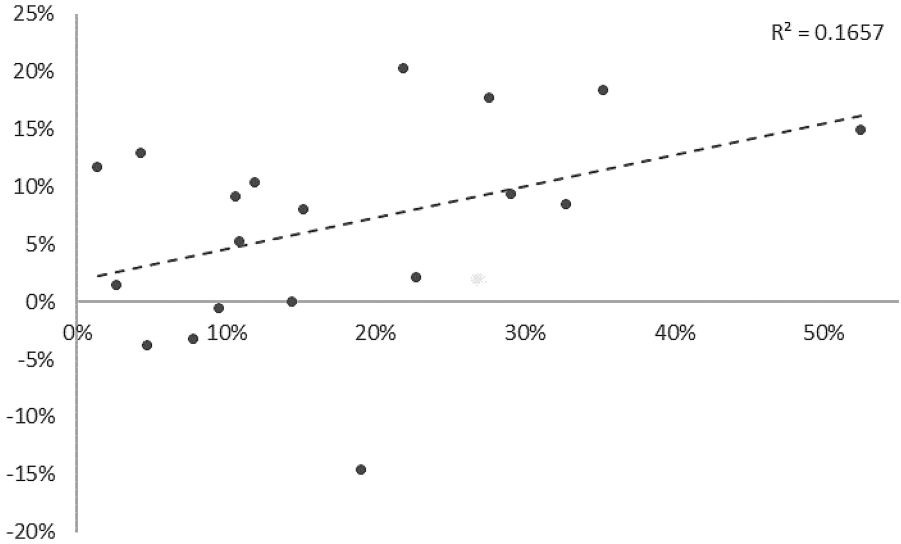
B. The Share of Religious Agencies in a State and Disparities in Treatment of Same-Sex and Opposite-Sex Couples

We explored whether states differed in their patterns of sexual orientation disparities based on the relative share of religious agencies operating within them. The conditions for conducting this analysis were not ideal, as many states operated only a handful of agencies (seventeen states had five or fewer agencies available for contact). This meant that our estimates of discrimination in these states, even with three waves of inquiry, were based on a very small number of observations. Furthermore, although our methodology aimed to track and include all agencies operating in all states, it is possible that for reasons unknown

to us we did not succeed in finding all such agencies, possibly skewing the results in states with a small number of agencies.

To narrow these concerns, we limited our analyses to states with 20 agencies or more (and therefore at least 60 observations across three waves of inquiry). We found a modest relationship (R^2=0.166) between the share of religious agencies in the state and the average gap in positive responses between opposite-sex and same-sex couples. These results suggest that *Fulton* may have had greater implications for states where religious agencies compose a large share of the foster care services market.

FIGURE A4. THE SHARE OF RELIGIOUS AGENCIES AND THE AVERAGE GAP IN POSITIVE RESPONSES BETWEEN OPPOSITE-SEX AND SAME-SEX COUPLES, ACROSS THE THREE WAVES, FOR STATES WITH MORE THAN TWENTY AGENCIES



Data is plotted across three waves, for states with more than twenty agencies. Each state is a dot.

C. *Demographic and Attitudinal Controls*

TABLE A3. AGREEMENT TO PROVIDE SERVICE, LEGAL JURISDICTION, RELIGIOSITY, AND ATTITUDES TOWARDS GAY COUPLES

	Agreement to Provide Service	
	(1)	(2)
Same sex	-0.114***	-0.106***
	(0.035)	(0.036)
Antidiscrimination Rule	-0.015	-0.047
	(0.040)	(0.056)
Religious Exemption Rule	0.020	0.063
	(0.048)	(0.057)
Same-Sex* Antidiscrimination Rules	0.085**	0.076*
	(0.039)	(0.040)
Same-Sex* Religious Exemption Rules	0.001	-0.007
	(0.047)	(0.047)
Antidiscrimination* Religious Exemption Rules	-0.054	-0.009
	(0.079)	(0.090)
Same-Sex*Antidiscrimination*Religious Exemption Rules	-0.151**	-0.143*
	(0.077)	(0.079)
Constant	0.616***	0.517
	(0.036)	(0.372)
Agency Fixed Effects	Yes	Yes
Agencies	1,234	1,200
Observations	2,468	2,400
Demographic and Attitudinal Controls	No	Yes

Note: ***Significant at the 1 percent level, **Significant at the 5 percent level, *Significant at the 10 percent level.

D. Email Templates

First wave

Hi,

My name is [name], and my [husband/wife] [spouse name] and I are considering to become foster parents. We understand that your agency can help us go through this process and would like to hear some more. I work as a part time soccer team coach and [spouse name] is a human resource manager. We are currently approaching our 40's and having met in high school, we have been together for quite a while now. Together, we have built a nice little family with two beautiful children.

Our main question regarding foster care and taking another child into our home is the response of our children. We saw some articles discussing this matter online, but we are not sure about their accuracy. We were wondering if we can talk to someone or whether there is an orientation that addresses these and other questions we might have?

Best wishes,

[name]

Second wave

Hello,

I am writing to inquire about your foster care program. If I am not mistaken, I understood from my searches that your organization can help with the certification required to become a foster home, for which my [husband/wife] [spouse name] and I wish to apply. We would like to have a chat with a representative, so if you could email us back with a relevant phone number and the best hours to call the office in the coming week, we would be happy to give you a call.

We assume some information about us will save some time in matching us with the relevant representative: My [husband/wife] [spouse name] works as project manager and I work as a speech therapist. We also have an incredible 14-year-old boy and a dog.

Ideally, we envision fostering a child under 12 years old, though we are open to discuss the option of an older child.

Thank you very much for the important work you do. We look forward to talking.

[name]

Third wave

Hello,

My name is [name], and I am contacting you to inquire about your foster care program. My [husband/wife], [spouse name], and I have been married for a little over a decade and for the past five years are raising our daughter together. I am 39 years old and [spouse name] is 37 years old. We are thrilled about the idea of fostering.

Which brings me to my question. We would love to get the number of a social worker or a representative whom we can ask some questions. We did come across your agency's general phone number online, but we would like to talk to someone who can explain to us the first steps of the process. Would you be able to provide us with a relevant phone number of such a representative?

Thanks, and have a good day,

[name]

Fourth wave

Hello,

I am contacting you to inquire about your foster care program. My [husband/wife], [spouse name], was inspired by an interview with one of your foster parents a while ago and we decided to reach out and learn more. Is there any orientation we can attend for people that are considering fostering? Zoom is preferable.

Thank you very much and have a good day,

[name]

E. Additional Information on Inquiry Emails

As previously mentioned, each agency received four inquiries throughout the experiment, from four different characters. Each inquiry differed stylistically to avoid raising suspicion on the part of agencies. Those differences included, for example, a different greeting, a different number of paragraphs, or a request (to receive the date of an upcoming orientation or a phone number of a representative). We also included different demographic data about the contacting family.

Since we used only eight characters to contact roughly 1,600 agencies each wave, each character sent hundreds of inquiries over a short period. To avoid being marked as spam by Gmail, in each wave we created several versions of the waves' inquiry letter, with minor changes between the versions. Each character used two of those versions and alternated between them so that every day a different version was sent.

Additionally, we randomized the timing of the emails, sending them in such a way as to ensure that foster care networks in the study with multiple branches would not receive all emails on the same day. This was done due to the concern that a network with multiple branches and unique branch emails might still be served by one server which could spot mass emails as spam.

We used the GMASS app to send the emails. GMASS alerted us when an email was blocked, and in that case, we sent emails manually one or two days after the wave was over. Due to the concern of being blocked by Gmail, all emails sent were timed with lags from one another. We sent them between 10 AM and 2 PM Eastern Time. Likewise, the online forms were filled out between 10 AM and midnight Eastern Time.

F. Public Agencies Sub-Database

The public agencies sub-database included two types: states with a single statewide contact and states with separate contacts for different counties or regions (for the sake of unified terminology, we call them "branches," similar to private networks that have different branches). States with a single statewide contact were states that indicated a single statewide contact for prospective foster parents—an email or an online form—as well as states that referred prospective parents to the counties/regions, but provided only phone numbers as a way to contact the regions (n = 21). Table A4 breaks down these distinctions.

TABLE A4. STATES WITHOUT BRANCH EMAILS

State with Email
Alaska
Arizona
Connecticut
Delaware
Hawaii
Idaho
Illinois
Mississippi
New Hampshire
North Carolina
Oregon
Rhode Island
South Carolina
Vermont
West Virginia
State with Online Form
Arkansas
Kentucky
Michigan
North Dakota
Tennessee
Utah

States with separate branch contacts included states that referred prospective parents to the counties/regions (without providing contact

information) where at least some of those local departments' contacts were found. Separate entries were entered for each local office, by region or county, depending on the state (Table A5).

TABLE A5. STATES WITH BRANCH CONTACT

State	Branches
Alabama	66
California	39 (13 with online contact forms only)
Colorado	44 (5 with online contact forms only)
Georgia	52
Indiana	18
Kansas	15
Louisiana	9
Massachusetts	5
Minnesota	74
Missouri	14
Montana	5
Nebraska	5
Nevada	2
New Jersey	2
New Mexico	9
New York	53 (3 with online contact forms only)
Ohio	80 (49 with online contact forms only)
Oklahoma	2
Pennsylvania	24 (2 with online contact forms only)
South Dakota	6
Virginia	22 (7 with online contact forms only)
Washington	6
Wisconsin	78
Wyoming	29 (all with online contact forms only)

Lastly, there were five states left out of the public agencies database in one of two cases: when state websites were not accessible (due to out-of-state location or repeated website failures; these were Florida and Maryland) or states whose website did not provide any contact information other than phone numbers (Texas, Iowa, and Maine).

G. Phone Number Properties

To increase the reliability of the profiles, we used an internet platform to create eight local phone numbers (four before *Fulton* and four after). The numbers were chosen from area codes of highly populated areas and were only noted in the signatures of the inquiry emails. That is, we did not invite or ask agencies to call us.

TABLE A6. PHONE NUMBER WAVE LOCATION

	1st and 2nd Wave Locations	**3rd and 4th Wave Locations**
John	California	Georgia
Robert	Florida	Illinois
Scott	Texas	Massachusetts
David	California	Pennsylvania

In the first wave, we used an automatic voicemail message that stated the first name of the character and allowed the caller to leave a message. We canceled this option in the following waves to preclude the option to leave a recorded message instead of replying over email (as email replies were our outcome variable). The first 271 voice messages we received in the first wave were untraceable to specific agencies, as callers often did not provide sufficient details that would enable identification of the agency from where they were calling.

Over the entire experiment, we received 57 SMS messages and 2,164 (unanswered) calls. To the extent we could identify the agency from which a contact was made, we also coded this information, but for the vast majority of cases this was not possible.

H. Bethany Christian Services Special Procedure

In the first wave we discovered that Bethany Christian Services, one of the largest foster care agencies in the United States, with dozens of branches across the country, uses a shared database of inquirers among its different branches. This allows representatives from the different branches to see if a character contacted another branch in different geographical locations, presenting a risk of contamination for our design. For this reason, following the first wave we made sure that a character would send inquiries to agencies that were geographically proximate to one another, assuming it was less suspicious for a couple to inquire with different branches in neighboring states than to inquire with branches located in states at different ends of the country. Since we already used the original eight characters to send emails to all parts of the country, we created eight new characters for Bethany. We then divided the branches of Bethany Christian Services into four groups based on geographic proximity. Two new characters were assigned to each group (one a same-sex and the other an opposite-sex couple) so that these two characters contacted branches in the same geographical area. Each one of the eight new characters sent inquiries that had the same properties of an original character, as listed below:

Table A7. Bethany Christian Services Characters

New Character	Spouse Name	Corresponding Character
Matthew Reed	Jennifer	John Peterson
Matthew Downy	Emily	Robert Hannagan
Jason Miller	Ashley	John Anderson
Jason Allen	Susan	Robert Wood
Michael Olson	Peter	Scott Turner
Michael Taylor	Andrew	David Howard
Christopher Mitchell	Kevin	Scott Harris
Christopher Hirst	Steven	David Stewart

I. Missing Observations and Exclusions

Exclusion of agencies occurred for several reasons. Most exclusions were due to missing observations that resulted from blocked emails or technical barriers to sending an email or submitting an online form in one or more waves (313 exclusions). In eighty-two cases, we removed the agency because it received more than one inquiry in the same

wave (such an event mostly occurred when our email was forwarded by one agency to another agency which was also in our database; the forwarding agency was retained in the sample). Twelve agencies were removed due to a mistaken reply by a member of the research team and five agencies were removed due to an expression of suspicion (before the fourth wave, see below). One agency was removed because its reply indicated that the representative did not understand that the sender was a same-sex couple. In total, 413 agencies were removed.

We focus on the analyses of agencies with complete observations, yet a wave-by-wave analysis that included all agencies with observations for that wave, even if they have less than three complete waves, yielded the same results (in this analysis we had 1311–35 agencies per wave).[259]

TABLE A8: THE FINAL SAMPLE BY AGENCY TYPE (ORIGINAL NUMBERS IN BRACKETS)

	Private Agencies		Public Agencies	Total
	Religious	**Non-Religious**		
Final N	230	581	424	1,235
(Original N)	(292)	(807)	(549)	(1,648)
Control N	38	114	68	220
(Original Control N)	(47)	(131)	(79)	(257)

J. Exclusion of Wave 4 Due to Suspicion

During the research, we encountered a few different indications that the fourth wave of inquiries raised suspicion among a non-negligible group of agencies. First, the virtual phone service we used to assign the profiles phone numbers canceled the service at the end of the fourth wave, stating suspicion of inauthentic activity. Second, we received a series of explicitly suspicious replies from agencies in that wave, after barely receiving any suspicious replies in previous waves. This was discovered in the coding process. Concomitantly, we discovered that the pattern of responses in the fourth wave was at odds with

[259] In the case of technical problems, we excluded only the observation of the wave where the problem occurred from the wave-by-wave analysis. If the problem could have induced suspicion (e.g., crossed inquiries, mistake, or explicit expression of suspicion), we excluded the problematic wave and any following waves.

prior waves. Whereas in previous waves same-sex couples consistently received fewer positive responses than opposite-sex couples, in the fourth wave same-sex couples received 4% more positive responses than opposite-sex couples, and among public agencies the gap was 14% in favor of same-sex couples. Notably, this pattern resulted from a sharp drop in positive responses to *opposite-sex* couples, and not from a sudden spike in positive responses to same-sex couples. We then discovered that this unlikely pattern was found to characterize suspicion among subjects in correspondence studies similar to ours, particularly in the context of sexual orientation discrimination where the design involved repeated measurement of subjects.[260] Altogether, these three independent indications led us to conclude that the data from the fourth wave was unreliable.

[260] *See* Catherine Balfe, Patrick Button, Mary Penn & David J. Schwegman, *Infrequent Identity Signals, Multiple Correspondence, and Detection Risks in Audit Correspondence Studies*, 35 Field Methods 3, 11–15 (2023).