

Online Appendix

Your Honor's Misdeeds: The Consequences of Judicial Scandal on Specific and Diffuse Support*

Joshua Boston[†] Benjamin J. Kassow[‡] Ali S. Masood[§] David R. Miller[¶]

Abstract

Legitimacy is a bulwark for courts; even when judges engage in controversial or disagreeable behavior, the public tends to acquiesce. Recent studies identify several threats to the legitimacy of courts, including polarization and attacks by political elites. We contribute to the scholarly discourse by exploring a previously unconsidered threat: scandal, or allegations of personal misbehavior. We argue that scandals can undermine confidence in judges as virtuous arbiters and erode broad public support for the courts. Using survey experiments, we draw on real-world judicial controversies to evaluate scandal's impact on specific support for judicial actors and their rulings and diffuse support for the judiciary. We demonstrate that scandals erode individual support but find no evidence that institutional support is diminished. These findings may ease normative concerns that isolated indiscretions by controversial jurists may deplete the vast "reservoir of goodwill" foundational to the courts.

* Authors are listed in alphabetical order per disciplinary convention, but this ordering does not reflect the authors' relative contribution to this work. We thank Eileen Braman, Kayla Canelo, Nathan Carrington, Jonathan King, Christopher Krewson, Elizabeth Lane, Jessica Schoenherr, Marcy Shieh, James Spriggs, Andrew Stone, Logan Strother, Guillermo Toral, and Nicholas Waterbury for helpful comments. We prepared this draft for presentation at the 2021 annual meeting of the American Political Science Association. An earlier iteration of this paper was presented at the 2021 annual meeting of the Midwest Political Science Association. This project adheres to the ethical and transparency obligations outlined in APSA's *Principles and Guidance for Human Subjects Research*. We provide details on adherence to these principles in Supplementary Information Section B. This research has been approved by the institutional review boards at Bowling Green State University (#1521845, Approved 01/17/2020; #1619591, Approved 7/21/2020; and #1638063, Approved 08/04/2020) and the University of North Dakota (#201912-138, Approved 12/18/2019). This research is supported by a grant from the Office of Sponsored Programs & Research (Grant # 33000321) and the Department of Political Science at Bowling Green State University.

[†] Assistant Professor, Department of Political Science, Bowling Green State University, Bowling Green, OH 43403; jbston@bgsu.edu

[‡] Associate Professor, Department of Political Science and Public Administration, University of North Dakota, Grand Forks, ND 58202; benjamin.kassow@und.edu

[§] Assistant Professor, Department of Politics, Oberlin College & Conservatory, Oberlin, OH 44074; amasood@oberlin.edu

[¶] Assistant Professor, Department of Political Science, International Affairs, and Public Administration, East Tennessee State University, Johnson City, TN 37604; millerdr3@etsu.edu

Appendix

Table of Contents

A	Examples of Judicial Scandal	SI.2
A.1	Ethical Scandals	SI.2
A.2	Sexual Scandals	SI.4
A.3	Financial Scandals	SI.5
B	Experiment Protocols and Materials	SI.6
B.1	Study 1: Supreme Court Nomination	SI.7
B.2	Study 2: Sitting Lower Court Judge	SI.9
B.3	Study 3: Court opinion in <i>Ramos v. Louisiana</i>	SI.12
B.4	Respondent Characteristics	SI.16
	Respondent Characteristics	SI.16
C	Empirical Analyses	SI.17
C.1	Robustness Checks	SI.18
C.2	Additional Randomizations	SI.18
C.3	Partisanship-Conditional Effects	SI.18
C.4	Interpreting Null Results	SI.19

List of Tables

SI.1	Respondent Descriptive Characteristics	SI.16
SI.2	Effect of Judicial Scandal on Specific and Diffuse Support (OLS, 0-1 Scales)	SI.22
SI.3	Effect of Judicial Scandal on Specific and Diffuse Support (OLS w/Covariates, 0-1 Scales)	SI.22
SI.4	Effect of Judicial Scandal on Specific Support (OLS, Ordinal Scales)	SI.23
SI.5	Effect of Judicial Scandal on Specific Support of Lower Court Judge (Multinomial Logistic Regression)	SI.23
SI.6	Effect of Judicial Scandal on Specific Support of Supreme Court Opinion (Multinomial Logistic Regression)	SI.24
SI.7	Effect of Judicial Scandal on Specific and Diffuse Support (OLS, 0-1 Scales, Including All Randomized Factors)	SI.24
SI.8	Effect of Judicial Scandal on Specific and Diffuse Support (OLS, 0-1 Scales, Copartisanship-Conditional)	SI.26
SI.9	Effect of Judicial Scandal on Specific and Diffuse Support (OLS, 0-1 Scales, Partisanship-Conditional)	SI.27
SI.10	Effect of Kavanaugh Authorship on Specific and Diffuse Support (OLS, 0-1 Scales, Knowledge-Conditional)	SI.28

A Examples of Judicial Scandal

Judicial scandals are not everyday occurrences in the United States, but allegations of judicial misbehavior surface from time to time, often attracting media attention and, in five instances in the post-World War II era, impeachment by the House of Representatives. In addition, similar allegations of judicial misconduct periodically arise in other countries around the world at the national level, exposing their respective courts to reputational harm. While no systematic accounting of judges' scandalous behavior exists, we present here examples of judicial scandals that illustrate the breadth of allegations judicial misconduct in both the United States and comparative contexts. Importantly, we used many of these examples to design the treatments in our two stylized experiments; in doing so, we promote the experiments' external validity by ensuring that the scandals we included in the vignettes mirrored the range of scandals to which the public has been exposed. To match the types of scandals we use as treatments (drawn from the classification of Basinger et al. 2013), we categorize these examples into ethical, sexual, and financial scandals.

A.1 Ethical Scandals

A.1.1 United States

- In 1968, Associate Justice Abe Fortas's nomination as Chief Justice was derailed when his acceptance of \$15,000 from clients of his former law firm to teach summer classes at American University came to light. A year later, *Life* reported that Fortas had been receiving a \$20,000 lifetime retainer from a former client recently convicted of financial crimes. Both incidents represented potential conflicts of interest, and Fortas resigned from the Supreme Court before impeachment proceedings began.¹
- In 1969, when being considered for the Supreme Court, Judge Clement Haynsworth of the Court of Appeals for the 4th Circuit was accused of having ruled in cases in which he had conflicts of interest. In one instance, he ruled in favor of a textile company that did business with Carolina Vend-a-Matic, a company in which he owned a one-seventh interest. His nomination ultimately failed in the Senate.²
- Judge Harry Claiborne of the District of Nevada was convicted of tax evasion in 1984 for failing to report over \$100,000 he earned from clients before ascending to the bench. Judge Claiborne refused to resign and intended to return to the bench after serving his 17 month prison sentence, but was impeached and removed from office in 1986.³
- In 1989, Judge Alcee Hastings of the Southern District of Florida was impeached and removed from the bench following allegations that he solicited a \$150,000 bribe from

¹ Pusey, Allen. "May 14, 1969: The Spectacular Fall of Abe Fortas." *ABA Journal*, April 1, 2020, <https://www.abajournal.com/magazine/article/the-spectacular-fall-of-abe-fortas>.

² "Clement F. Haynsworth Jr.; Judge Was Rejected as 1969 Supreme Court Choice." *Los Angeles Times*, November 23, 1989, <https://www.latimes.com/archives/la-xpm-1989-11-23-mn-3-story.html>.

³ "Harry Claiborne, 86, Is Dead; Was Removed as U.S. Judge." *The New York Times*, January 22, 2004, <https://www.nytimes.com/2004/01/22/us/harry-claiborne-86-is-dead-was-removed-as-us-judge.html>.

defendants in a case before him despite an acquittal for similar charges in a criminal trial. Hastings was elected to Congress in 1993 and still serves as a House member.⁴

- Judge Walter Nixon of the Southern District of Mississippi was impeached and removed from the bench in 1989 for committing perjury by providing false statements to a grand jury about his interceding in a state criminal case.⁵
- In 1989, Judge Robert Aguilar of the Northern District of California was convicted for his role in a conspiracy to influence the outcome of a case against a San Francisco mobster but resisted calls for his resignation and remained on the bench.⁶ Judge Aguilar’s conviction was eventually overturned on appeal in 1996, whereafter he agreed to retire in exchange for the federal government not retrying the case.⁷
- Judge Robert Collins of the Eastern District of Louisiana resigned from office in 1993 facing imminent impeachment proceedings following his criminal conviction for accepting a bribe in exchange for issuing a lenient sentence in a drug smuggling case.⁸
- In 2010, Judge Thomas Porteous of the Eastern District of Louisiana was impeached and removed from the bench over allegations that he received bribes from lawyers who argued cases before him.⁹
- In 2010, Judge Beverly Martin of the Court of Appeals for the 11th Circuit ruled in favor of an insurance company represented by her husband’s law firm. After the case was decided, Judge Martin recused herself and the case was reheard.¹⁰
- In 2018, Chief Judge Christopher Conner of the District Court for the Middle District of Pennsylvania presided over two cases where one of the participating law firms had recently hired his son. Chief Judge Conner ultimately recused himself from the cases.¹¹

⁴ Marcus, Ruth. “Senate Removes Hastings.” *Washington Post*, October 21, 1989, <https://www.washingtonpost.com/wp-srv/politics/campaigns/junkie/links/hastings102189.htm>.

⁵ Lewis, Neil A. “Senate Convicts U.S. Judge, Removing Him From Bench.” *The New York Times*, November 4, 1989, <https://www.nytimes.com/1989/11/04/us/senate-convicts-us-judge-removing-him-from-bench.html>.

⁶ MacLean, Pamela. “Federal Judge Sentenced to Six months in Prison.” *UPI*, November 1, 1990. <https://www.upi.com/Archives/1990/11/01/Federal-judge-sentenced-to-six-months-in-prison/7596657435600/>.

⁷ Holding, Reynolds. “Appeals Court Reverses Judge Aguilar’s Conviction.” *San Francisco Chronicle*, January 26, 1996. <https://www.sfgate.com/news/article/Appeals-Court-Reverses-Judge-Aguilar-s-Conviction-2996703.php>.

⁸ McQuaid, John. “Collins Resigns Federal Judgeship; Resignation Letter is Given to Clinton.” *The Times-Picayune*, August 7, 1993, Metro Section, Page 1.

⁹ Steinhauer, Jennifer. “Senate, for Just the 8th Time, Votes to Oust a Federal Judge.” *The New York Times*, December 8, 2010, <https://www.nytimes.com/2010/12/09/us/politics/09judge.html>.

¹⁰ O’Brien, Reity, Weir, Kytja, and Young, Chris. “Federal Judges Plead Guilty.” *Center for Public Integrity*, April 28, 2014, <https://publicintegrity.org/politics/federal-judges-plead-guilty>

¹¹ Fernandez, Bob. “Hershey School’s Law Firm Was Arguing Two Cases Before a Federal Judge—and Then Hired His Son.” *The Philadelphia Inquirer*, February 4, 2019, <https://www.inquirer.com/business/hershey-trusthires-judge-son-greenleaf-elliott-20190204.html>.

A.1.2 Comparative Contexts

- In 2017, Wael Shalaby, a deputy chief justice in the Egypt’s national administrative courts system, was arrested for having taken bribes. Shalaby committed suicide in prison shortly after being taken into custody.¹²
- In 2019, *The Intercept* released documents showing that Minister of Justice and Public Security Sergio Moro colluded with prosecutors to find evidence and discuss strategy in a corruption case against former Brazilian President Luiz Inácio Lula da Silva over which Moro had presided as a federal judge.¹³ Brazil’s Supreme Court later deemed Moro’s collusion to have unfairly biased the case and overturned da Silva’s conviction.
- In 2020, a judge on South Africa’s Western Cape High Court accused the court’s Judge President, John Hlophe, of indiscretions including attempts to influence which judges were assigned to a lawsuit filed by several organizations that sought to block then-President Jacob Zuma’s acquisition of a nuclear power plant from Russia.¹⁴

A.2 Sexual Scandals

A.2.1 United States

- During his 1991 confirmation process, Justice Clarence Thomas faced allegations from Anita Hill that he repeatedly sexually harassed her while serving as a political appointee in two federal agencies. After three days of contentious nationally-televised hearings, the Senate ultimately confirmed Thomas to the Supreme Court.¹⁵
- Judge Samuel Kent of the Southern District of Texas resigned from the bench in 2009 after the House impeached him, but before the Senate convicted him, following allegations of sexually assaulting multiple courthouse employees.¹⁶
- In 2010, Judge Jack Camp of the Northern District of Georgia retired after being caught in a Federal Bureau of Investigation sting operation for purchasing and using drugs with an exotic dancer.¹⁷

¹² “Egyptian judge facing corruption charge hangs himself: lawyer.” *Reuters*, January 2, 2017, <https://www.reuters.com/article/us-egypt-corruption/egyptian-judge-facing-corruption-charge-hangs-himself-lawyer-idUSKBN14MONJ>.

¹³ Fishman, Andrew, Martins, Rafael Moro, Demori, Leandro, de Santi, Alexandre, and Greenwald, Glenn. “Breach of Ethics.” *The Intercept*, June 9, 2019, <https://theintercept.com/2019/06/09/brazil-lula-operation-car-wash-sergio-moro/>.

¹⁴ Thamm, Marianne. “Sex, lies, physical assault & court rigging — all in a day’s work for John Hlophe, claims his deputy.” *Daily Maverick*, January 22, 2020, <https://www.dailymaverick.co.za/article/2020-01-22-sex-lies-physical-assault-court-rigging-all-in-a-days-work-for-john-hlophe-claims-his-deputy/>.

¹⁵ Jacobs, Julia. “Anita Hill’s Testimony and Other Key Moments From the Clarence Thomas Hearings.” *The New York Times*, September 20, 2018, <https://www.nytimes.com/2018/09/20/us/politics/anita-hill-testimony-clarence-thomas.html>.

¹⁶ Neil, Martha. “Federal Judge Samuel Kent Resigns, as Senate Impeachment Trial Looms.” *ABA Journal*, June 25, 2009, https://www.abajournal.com/news/article/federal_judge_samuel_kent_resigns_as_senate_impeachment_trial_looms.

¹⁷ Rankin, Bill. “Ex-Judge Camp Sentenced to 30 Days in Prison.” *Atlanta Journal-Constitution*, August 11, 2012. <https://www.ajc.com/news/local/judge-camp-sentenced-days->

- In 2015, Judge Mark Fuller of the Middle District of Alabama resigned from the bench after reaching a plea deal concerning domestic violence charges.¹⁸
- In 2017, Judge Alex Kozinski of the Court of Appeals for the 9th Circuit resigned after several former female staffers alleged that he had sexually harassed them.¹⁹
- During his 2018 confirmation proceedings, Justice Brett Kavanaugh faced multiple allegations of sexual misconduct, most notably from Dr. Christine Blasey Ford concerning sexual assault at a high school party. Despite the accusations, Kavanaugh was confirmed by the Senate in a narrow 50-48 vote.²⁰

A.2.2 Comparative Contexts

- In 2020, an internal investigation by the High Court of Australia concluded that former Justice Dyson Heydon had sexually harassed at least six women on his staff during his ten years on the court (2003-2013).²¹
- In 2021, nude pictures and salacious text messages exchanged by High Court judge Justice Thompson Mabhikwa of Zimbabwe and a member of his court’s staff were publicly leaked. Many Zimbabwean lawyers and court observers argued that the relationship with a subordinate violated judicial ethics rules and called on the judge to resign. As of this writing, Justice Thompson Mabhikwa remains on the High Court.²²

A.3 Financial Scandals

A.3.1 United States

- Shortly before nominating Stephen Breyer in 1993, the Clinton administration discovered that Breyer had not paid Social Security taxes for a part-time housekeeper his family had employed for 13 years.²³
- In 2016, shortly after her retirement from the United States Tax Court, Diane Kroupa was indicted for conspiracy to defraud the United States for intentionally understating

prison/FQhgyRbi1JD1oK28fQRGoJ/.

¹⁸ Troyan, Mary. “U.S. District Judge Mark Fuller will resign.” *Montgomery Advertiser*, May 29, 2015, <https://www.montgomeryadvertiser.com/story/news/2015/05/29/us-district-judge-mark-fuller-will-resign/28178951/>.

¹⁹ Zaposky, Matt. “Prominent Appeals Court Judge Alex Kozinski Accused of Sexual Misconduct.” *Washington Post*, December 8, 2017, https://www.washingtonpost.com/world/national-security/prominent-appeals-court-judge-alex-kozinski-accused-of-sexual-misconduct/2017/12/08/1763e2b8-d913-11e7-a841-2066faf731ef_story.html

²⁰ Silverstein, Jason. “Brett Kavanaugh confirmed to Supreme Court by smallest margin since 1881.” *CBS News*, October 6, 2018, <https://www.cbsnews.com/news/brett-kavanaugh-confirmed-to-supreme-court-by-smallest-margin-in-modern-history/>.

²¹ Cave, Damien and Kwai, Isabella. “A Sexual Harasser Spent Years on Australia’s Top Court, an Inquiry Finds.” *The New York Times*, June 23, 2020. <https://www.nytimes.com/2020/06/23/world/australia/dyson-heydon-high-court-metoo.html?smid=url-share>.

²² Munyoro, Fidelis. “Just in: Judge under pressure to resign.” *The Herald*, February 26, 2021, <https://www.herald.co.zw/just-in-judge-under-pressure-to-resign/>.

²³ Berke, Richard L. “Favorite for High Court Failed to Pay Maid’s Taxes.” *The New York Times*, June 13, 1993, Section 1, Page 1.

her and her husband’s taxes owed to the government by \$450,000. The following year, Kroupa and her husband were convicted and sent to federal prison.²⁴

A.3.2 Comparative Contexts

- In 1993, Justice Veeraswami Ramaswami of the Supreme Court of India faced removal proceedings amid allegations that he misused public funds for personal purposes, such as an extravagant renovation of his official residence. Despite an official report confirming many of the charges, the removal motion failed in India’s parliament.²⁵

B Experiment Protocols and Materials

In this section, we describe the protocols, provide the vignette and question wordings for each of our experiments, and detail the descriptive statistics of our survey samples.

Our three experiments utilized two different respondent pools—Amazon’s Mechanical Turk, or MTurk (Study 1), and Lucid (Studies 2 and 3). Both pools are commonly used in political science research but have different recruitment and sampling methods. On MTurk, researchers solicit workers to complete their surveys as Human Intelligence Tasks (HITs), and their survey samples consist of those workers who chose to complete the surveys (Berinsky, Huber, and Lenz 2012). Alternatively, Lucid recruits panels of survey-takers and provides researchers with survey samples that are representative of the US public for common demographic characteristics such as race, gender, and party identification (Coppock and McClellan 2019). Thus, MTurk is a convenience sample while Lucid is a representative sample. We present the demographic characteristics of our three survey samples in Table SI.1 at the end of Supplemental Information Section B.

Survey samples from MTurk are generally suitable for making inferences about the political behavior of the US population at large (i.e., sample average treatment effects \approx population average treatment effects; see Coppock, Leeper, and Mullinix 2018; Coppock 2019), and the substantive similarity of our findings from our MTurk and Lucid experiments (Study 1 versus Studies 2 and 3, respectively) suggest that the treatment effects from our MTurk sample generalize to the US population. Slight differences in the magnitudes of our treatment effects for specific support in Studies 1 and 2 might be attributable to the differences in samples; namely, that the sexual scandal sparked a negative effect for specific support that is distinguishably larger than the other types of scandal in Study 1, but not in Study 2, might result from MTurk samples tending to include more younger, liberal, and Democratic respondents who are more likely to embrace the #MeToo movement and punish political

²⁴ “U.S. Tax Court judge, husband from Minnesota sent to prison for claiming 9 vacations as ‘business expenses’.” *Grand Forks Herald*, June 22, 2017, <https://www.grandforksherald.com/4287735-us-tax-court-judge-husband-minnesota-sent-prison-claiming-9-vacations>.

²⁵ Ghose, Sanjoy. “Carpets, Bedsheets, Towels and Intrigue: The Story of Justice V. Ramaswami’s Impeachment.” *The Wire*, July 22, 2020, <https://thewire.in/law/justice-v-ramaswami-impeachment-lok-sabha-kapil-sibal-supreme-court>.

figures for sexual misconduct more strongly (Costa et al. 2020; Holman and Kalmoe 2021).²⁶ However, that all types of scandal employed in Studies 1 and 2 are associated with statistically distinguishable and negative effects on specific support, and that our results for diffuse support in Studies 1 and 2 are roughly similar, suggests that the differences in samples do not drive any substantive conclusions we draw.

Finally, while Study 3 leverages the real-world political milieu by incorporating real Supreme Court justices and a real Supreme Court case, Studies 1 and 2 utilize hypothetical federal judges. We opt for hypothetical judges in these studies because while we need to hold all elements of the vignettes constant besides allegations of ethical, financial, and sexual scandals to maintain internal validity, we cannot credibly tell respondents that any single real federal judge has been accused of all of these types of scandals. Our use of hypothetical judges mirrors the approach of many other published survey experiments on public attitudes towards the courts (Gibson and Caldeira 2013; Nelson 2018; Ono and Zilis 2022). While using hypothetical judges sacrifices some degree of external validity, Brutger et al. (2022) demonstrate that survey experiments using abstracted details, such as hypothetical political actors, recover similar treatment effects as otherwise identical experiments using real-world stimuli. Additionally, while it is possible that some respondents may have known that the judges we mentioned in our Study 1 and Study 2 vignettes did not actually serve as federal judges, the public’s generally low knowledge of the federal courts and the judges serving on them makes this possibility unlikely.²⁷

B.1 Study 1: Supreme Court Nomination

We fielded our Supreme Court nomination experiment on MTurk in January 2020 with approximately 1400 US respondents.²⁸ Before proceeding to assigned vignette, respondents completed a battery of demographic questions and two-attention check tasks drawn from (Berinsky, Margolis, and Sances 2014).²⁹

All respondents were presented with a vignette consisting of a stylized account of a sitting federal judge’s nomination to the Supreme Court. The common content of the vignette for all respondents consisted of details about the judge’s background, a discussion of the Senate’s

²⁶ As Coppock (2019) details, convenience samples such as MTurk yield treatment effects substantively different from those observed in the target population (here, the US population) if treatments engender heterogeneous effects among respondents and the distribution of the respondent characteristics connected to those heterogeneous effects differs from that in the target population Coppock, Leeper, and Mullinix (see also 2018). Since the MTurk sample might contain more respondents who react more negatively to sexual misconduct, the associated treatment effect might be larger than in the target population.

²⁷ In a March 2022 C-SPAN/Pierrepoint poll, fewer than half of survey respondents could name at least one sitting Supreme Court justice (“Supreme Court Survey,” *C-SPAN*, March 2022, <https://static.c-span.org/assets/documents/scotusSurvey/fullSurvey.2022.b.pdf>).

²⁸ Following Kennedy et al. (2020), we screened respondents’ locations using their IP addresses and did not allow persons located outside of the US or using a VPN to participate.

²⁹ The attention tasks asked respondents to indicate their commonly-consumed news sources and any emotions they were feeling. In both questions, we embedded instructions to ignore the question and choose one or more specific answers to show that they were paying attention. We used the same two tasks in each of our three experiments.

vetting of the judge, and an indication that the judge was ultimately confirmed. While the fact that these accounts are stylized may seem troublesome to some individuals, in fact, given what we know of the American public, we believe this is unlikely to be a significant issue. After all, several studies have showed that only somewhere between one-half and two-thirds of Americans can name a *single* U.S. Supreme Court justice, let alone all nine—a C-SPAN study of “likely voters” in 2018 found that 52 percent of likely voters could name a single U.S. Supreme Court justice. Even more surprisingly, a survey commissioned by FindLaw from 2012 found that two-thirds of surveyed American adults could not name a single U.S. Supreme Court justice, and that only one percent of Americans could name all nine justices correctly).

Two facets of the vignette were randomized across respondents. First, respondents were randomly assigned to receive information about scandalous accusations made against the judge during the Senate’s vetting. While respondents in the control condition received no additional information for this facet, respondents in the scandal conditions received information that the judge had been implicated in an ethical, sexual, or financial scandal. Second, to account for potential copartisanship-conditional effects (Bartels and Johnston 2013), respondents were randomized to be told that the judge had been nominated to the Supreme Court by Republican President George W. Bush or Democratic President Barack Obama.

After reading their assigned vignettes, respondents were asked to indicate their specific support for the judge confirmed to the Court on a four-point scale and their diffuse support for the Court using the six statements developed by Gibson, Caldeira, and Spence (2003).

Upon completing the survey, we debriefed respondents by explaining that our vignette described a hypothetical scenario and that the featured judge never served as a federal judge or was nominated to the Court.³⁰ Respondents received \$0.35 for completing the survey.

B.1.1 Vignette

Several years ago, [*nominating president*] nominated John Clark to serve as a justice on the United States Supreme Court. Judge Clark was 47 years old at the time of nomination and had served as a judge on the US Court of Appeals for the Sixth Circuit for 10 years. Judge Clark received his law degree from Yale Law School and lived in Cincinnati, Ohio, at the time of nomination with his wife and their two children. When Judge Clark’s nomination was announced, many of his colleagues and several legal organizations released statements praising his legal skills and supporting his nomination.

The Senate Judiciary Committee vetted Judge Clark’s professional and personal background and questioned him in a televised hearing. [*scandal*] In the end, the US Senate approved Judge Clark’s nomination, and he assumed his role on the US Supreme Court.

- **Randomizations**

³⁰ This mild deception was necessary because judicial scandals are sufficiently infrequent to provide counterfactual scenarios (i.e., instances in which two judicial nominations were identical save scandal).

- Nominating president
 - * Democratic President Barack Obama
 - * Republican President George W. Bush
- Scandal
 - * *Blank*
 - * During the vetting process, reports emerged that Judge Clark sexually harassed several women who work at the courthouse where he serves.
 - * During the vetting process, reports emerged that Judge Clark violated judicial ethics guidelines by issuing rulings in several cases where one of the litigants was represented by his brother’s law firm.
 - * During the vetting process, reports emerged that Judge Clark failed to pay federal taxes for a housekeeper his family employed for several years.

B.1.2 Question Wordings

Do you approve or disapprove of John Clark serving as a justice on the US Supreme Court?

- Strongly approve
- Somewhat approve
- Somewhat disapprove
- Strongly disapprove

Do you agree or disagree with each of the following statements? (*Response options for all questions: Strongly agree, somewhat agree, somewhat disagree, strongly disagree*)

- If the US Supreme Court started making a lot of decisions that most people disagree with, it might be better to do away with the Supreme Court altogether.
- The right of the Supreme Court to decide certain types of controversial issues should be reduced.
- The Supreme Court can usually be trusted to make decisions that are right for the country as a whole.
- The decisions of the US Supreme Court favor some groups more than others.
- The US Supreme Court gets too mixed up in politics.
- The US Supreme Court should have the right to say what the Constitution means, even when the majority of the people disagree with the Court’s decisions.

B.2 Study 2: Sitting Lower Court Judge

We fielded our lower court judge experiment in August 2020 using Lucid Theorem with approximately 1650 US respondents. Before proceeding to their assigned vignette, respondents completed a battery of demographic questions and two attention check tasks drawn from (Berinsky, Margolis, and Sances 2014).

All respondents were presented with a stylized news article about a sitting judge on the US Court of Appeals for the District of Columbia who pledges to remain on the bench despite personal news that might prompt resignation or retirement.

We randomized the same two facets of the vignette as we randomized in Study 1. First, respondents were randomly assigned to receive information about the reason why the judge’s future tenure is in doubt. We inform respondents in the control condition that the judge is undergoing treatment for adrenal cancer, while respondents in the scandal conditions learn that the judge has been implicated in an ethical, sexual, or financial scandal. Second, to account for potential copartisanship-conditional effects, respondents were again randomized to receive information that the judge had been nominated to the Supreme Court by Republican President George W. Bush or Democratic President Barack Obama.

After reading their assigned vignettes, respondents were asked to indicate their specific support for the judge on a four-point scale and their diffuse support for the federal courts using the six statements developed by Gibson, Caldeira, and Spence (2003).³¹

Upon completing the survey, we debriefed respondents by explaining that our vignette described a hypothetical scenario and that the featured judge (Riley Emerson) never served as a federal judge.³² We did not compensate respondents directly, but Lucid provides panelists with undisclosed monetary incentives for completing surveys on their platform.

B.2.1 Vignette

Federal Judge Emerson Vows to Remain on the Bench After [headline]

The Associated Press

Riley Emerson, a high-profile federal appellate court judge in Washington D.C., is to remain on the federal bench despite [scandal].

In a statement on Monday, Judge Emerson, an appointee of President [nominating president], said [remaining] was important for the continuity of justice at the D.C. Circuit Court—widely considered the second most important court of the United States.

“I am confident that [self-defense],” Emerson said in the statement. “I am proud of serving the public in the courtroom. So, I am making the decision to remain working with my fellow, committed federal judges.”

Judge Emerson, a graduate of Harvard University and Yale Law School, has served on the United States Court of Appeals for the D.C. Circuit for over a decade.

- **Randomizations**

- Headline

- * Announcing Cancer Diagnosis

³¹ Because the diffuse support question battery used by Gibson, Caldeira, and Spence (2003) focused exclusively on the Supreme Court, we made slight modifications to the question wordings to encompass the full federal judiciary.

³² This mild deception was necessary because while judicial scandals occur, they are sufficiently infrequent to provide counterfactual scenarios (i.e., instances in which two federal judges are identical save scandal).

- * Sexual Harassment Allegations
- * Ethics Violations Allegations
- * Tax Violations Allegations
- Scandal
 - * being diagnosed with adrenal cancer.
 - * accusations of sexual harassment. Last Thursday, reports emerged that Judge Emerson made unwanted sexual advances towards several law clerks at the courthouse where he works.
 - * accusations of ethics violations. Last Thursday, reports emerged that Judge Emerson violated ethics guidelines by ruling on several cases where one of the litigants was represented by a law firm his brother works for.
 - * accusations of tax fraud. Last Thursday, reports emerged that Judge Emerson failed to pay federal taxes for a housekeeper Emerson employed for several years.
- Nominating President
 - * George W. Bush
 - * Barack Obama
- Remaining
 - * family and friends had urged him to remain and defend himself, and that doing so
 - * he has been working with a medical team for months, but that remaining on the bench
- Self-Defense
 - * I will be cleared of any wrongdoing
 - * my medical team will help me overcome this disease

B.2.2 Question Wordings

Do you approve or disapprove of Riley Emerson serving as a judge on the DC Circuit Court?

- Strongly approve
- Somewhat approve
- Somewhat disapprove
- Strongly disapprove
- Don't know

Do you agree or disagree with each of the following statements? (*Response options for all questions: Strongly agree, somewhat agree, somewhat disagree, strongly disagree, don't know*)

- If the federal courts started making a lot of decisions that most people disagree with, it might be better to do away with the federal courts altogether.
- The right of the federal courts to decide certain types of controversial issues should be reduced.

- The federal courts can usually be trusted to make decisions that are right for the country as a whole.
 - The decisions of the federal courts favor some groups more than others.
 - The federal courts gets too mixed up in politics.
 - The federal courts should have the right to say what the Constitution means, even when the majority of the people disagree with the federal courts' decisions.
-

B.3 Study 3: Court opinion in *Ramos v. Louisiana*

We fielded our Supreme Court opinion experiment in August 2020 using Lucid Theorem with approximately 850 US respondents. Before proceeding to their assigned vignette, respondents completed a battery of demographic questions and two attention check tasks drawn from (Berinsky, Margolis, and Sances 2014).

All respondents were presented with a stylized news article describing the Supreme Court's ruling in *Ramos v. Louisiana*, which the Court had decided in April 2020. In *Ramos v. Louisiana*, the Supreme Court incorporated to the states the Sixth Amendment's requirement that guilty verdicts for criminal trials must be unanimous. We use this case because it (1) was decided in the latter half of the Court's term, (2) dealt with a legal issue that is easy for the public to understand and not strongly polarized, and (3) included both Justices Gorsuch and Kavanaugh in the majority. This last feature enables us to plausibly attribute the majority opinion to either justice without loss of external validity. In fact, Gorsuch authored the majority opinion, which Kavanaugh joined; our debriefing materials informed respondents of the true opinion author.

We again randomized two facets of the vignette. First, respondents were randomly assigned to be told that the Court's opinion in the case was authored by Neil Gorsuch (control condition) or Brett Kavanaugh (scandal condition). Second, to account for potential moderating effects of the size of the majority coalition on the effect of scandal on public opinion, we also independently randomized whether the Court's decision in the case was unanimous (9-0) or divided (6-3).

After reading their assigned vignettes, respondents were asked to indicate their specific support for the Court's opinion on a four-point scale and their diffuse support for the federal courts using the six statements developed by Gibson, Caldeira, and Spence (2003).

Upon completing the survey, we debriefed respondents by providing the true opinion author and coalition size.³³ We did not compensate respondents directly, but Lucid provides panelists with undisclosed monetary incentives for completing surveys on their platform.

³³ This mild deception was necessary because while judicial scandals occur, they are sufficiently infrequent to provide counterfactual scenarios (i.e., instances in which two Supreme Court justices are identical save scandal and offer identical opinions in identical cases).

B.3.1 Design Considerations

In this experiment, we seek to examine how associating a justice who experienced high-profile allegations of scandal with a Supreme Court opinion affects the public’s specific support for that decision and its diffuse support for the Court as compared to associating a justice who has not experienced allegations of scandal with the same opinion. Thus, our treatment seeks to make salient the real-world presence and legal decision-making power of a scandalized justice. As we explain here, this experiment provides heightened external validity because it takes full advantage of the environment in which respondents experience the political world and incorporates respondents’ exposure to scandal two years prior to examine potential long-term effects.

We use Justice Brett Kavanaugh, who was accused of sexual assault during his confirmation process and faced an additional day of hearings before the Senate Judiciary Committee to respond to those accusations, as our scandalized justice. In using real-world political actors as our treatments, it is impossible to have a precise counterfactual Justice Kavanaugh’ who did not face sexual assault allegations. However, we believe Justice Neil Gorsuch, who we designate as our control, is a reasonable counterfactual because he shares with Justice Kavanaugh the many characteristics of Supreme Court justices that are salient to the general public but has not previously faced allegations of misconduct; both were appointed by President Donald Trump, have conservative jurisprudential tendencies, are of the same race and gender, and are similar in age. Given these commonalities, any differences in specific and diffuse support we detect when the opinion in *Ramos v. Louisiana* is attributed to Justices Gorsuch or Kavanaugh should be attributable to the scandal faced by Kavanaugh.³⁴

To heighten external validity, we style our vignette after contemporaneous media reports of the same case. Accordingly, the vignette focuses on the case itself and mentions the justices themselves only to identify the positions they took on the case and the rationales for those positions. This focus on the case without reference to personal details about the justices themselves follows the style of accounts of Court decisions published by media outlets that routinely cover the Court.³⁵ By modeling our vignette on the type of news stories about the Court’s activities that typically appear in the media, our design examines whether making the presence of a scandalized justice salient in routine reporting affects public attitudes.

³⁴ In *Ramos v. Louisiana*, the Supreme Court incorporated to the states the Sixth Amendment’s requirement that guilty verdicts for criminal trials must be unanimous. We use this case because it was decided in the latter half of the Court’s term, dealt with a legal issue that is easy for the public to understand and not strongly polarized, and included both Justices Gorsuch and Kavanaugh in the majority. This last feature enables us to plausibly attribute the majority opinion to either justice without loss of external validity. In fact, Gorsuch authored the majority opinion, which Kavanaugh joined; our debriefing materials informed respondents of the true opinion author.

³⁵ For instance, *The New York Times*’ article detailing the Court’s ruling in *Ramos v. Louisiana* does not mention individual justices other than to identify their votes and quote from their opinions (Liptak, Adam. “Supreme Court Bans Non-Unanimous Jury Verdicts for Serious Crimes.” *The New York Times*, April 20, 2020, <https://www.nytimes.com/2020/04/20/us/politics/supreme-court-unanimous-verdicts.html>). Perusal of similar stories by *The New York Times* and the *Washington Post* on other Court decisions reveals that reporters rarely include contextual information about the justices when reporting on decisions.

Our design does not explicitly remind respondents of the sexual assault allegations made against Justice Kavanaugh in his confirmation hearings, but rather assesses whether those allegations exert influence on respondents' evaluations of the Court a year and a half into his tenure. Public opinion polls fielded during Justice Kavanaugh's confirmation and his tenure on the Court suggest that the allegations made against him had lasting effects on public attitudes. A YouGov/Economist poll fielded September 30-October 2, 2018—shortly before the Senate's confirmation vote on October 6, 2018—found that 60% of respondents reported having “heard a lot” about the nomination; thus, were the sexual assault allegations made against Kavanaugh to have effects on public support for the judiciary, a majority of the public were likely aware of those allegations and could draw upon those considerations when evaluating the Court in the future.³⁶ A more recent YouGov/Economist poll fielded in March 2021 which asked members of the public to indicate their favorability towards each Supreme Court justice indicates that Justice Kavanaugh is relatively well-known among his colleagues, as only 34% of respondents declined to offer an opinion about him—the second-lowest share of “don't know” answers among all justices, second only to Clarence Thomas at 33%.³⁷ In addition, Justice Kavanaugh received the highest proportion of “unfavorable” opinions among all justices (35%), suggesting that some of the reputational damage he incurred from the sexual assault allegations made during his confirmation hearings may linger in the public's mind two and a half years later. Thus, if judicial scandal exerts long-term effects on public attitudes towards the Court, respondents likely had sufficient prior exposure to and lasting recollections of the allegations made against Justice Kavanaugh to be able to draw on those considerations when expressing their opinions about the Court.

Some may suggest that the absence of scandalous allegations in the vignette makes our design weak, but we argue that this design makes our experiment externally valid because it leverages real-world variation in the scandalized status of justices, mirrors the ways in which the public most often encounters information about the Court, and probes for the long-term effects of judicial scandal on specific and diffuse support which are most of interest to researchers. While an alternative design that explicitly mentions the allegations made against Justice Kavanaugh might uncover negative effects, this design would be of limited value because it makes salient judicial scandal in an artificial manner that does not mirror the real-world processes by which members of the public encounter news about the Court.

B.3.2 Vignette

Justice [Gorsuch/Kavanaugh] Leads [Unanimous/Divided] Supreme Court in Sixth Amendment Dispute

The Associated Press

³⁶ https://d25d2506sfb94s.cloudfront.net/cumulus_uploads/document/sodq3s3gan/econTabReport.pdf

³⁷ Frankovic, Kathy. “How well-known are the Supreme Court Justices?” *YouGov*, March 17, 2021, <https://today.yougov.com/topics/politics/articles-reports/2021/03/17/how-well-known-are-supreme-court-justices>.

WASHINGTON — In a recent decision, the Supreme Court ruled that the Constitution’s right to jury trial requires a unanimous verdict in state criminal trials. Justice [Neil M. Gorsuch/Brett M. Kavanaugh] wrote the Court’s opinion on behalf of a [9-0/6-3] majority.

The case, *Ramos v. Louisiana*, involves a defendant, Evangelisto Ramos, who was charged with second-degree murder and exercised his right to a jury trial. After deliberating, ten of the twelve jurors found that the prosecution had proven its case against Ramos beyond a reasonable doubt.

Under Louisiana’s non-unanimous jury verdict law, agreement of only ten jurors is sufficient to enter a guilty verdict, so Ramos was sentenced to life in prison without the possibility of parole.

In overturning Ramos’s conviction, Justice [Gorsuch/Kavanaugh] wrote that Louisiana did not secure “conviction constitutionally under the Sixth Amendment.”

“At the time of the Sixth Amendment’s adoption, the right to trial by jury included a right to a unanimous verdict,” Justice [Gorsuch/Kavanaugh] wrote. “When the American people chose to enshrine that right in the Constitution, they weren’t suggesting fruitful topics for future cost-benefit analyses. They were seeking to ensure that their children’s children would enjoy the same hard-won liberty they enjoyed.”

B.3.3 Question Wording

Do you approve or disapprove of the Supreme Court’s decision in *Ramos v. Louisiana*?

- Strongly approve
- Somewhat approve
- Somewhat disapprove
- Strongly disapprove
- Don’t know

Do you agree or disagree with each of the following statements? (*Response options for all questions: Strongly agree, somewhat agree, somewhat disagree, strongly disagree, don’t know*)

- If the US Supreme Court started making a lot of decisions that most people disagree with, it might be better to do away with the Supreme Court altogether.
- The right of the Supreme Court to decide certain types of controversial issues should be reduced.
- The Supreme Court can usually be trusted to make decisions that are right for the country as a whole.
- The decisions of the US Supreme Court favor some groups more than others.
- The US Supreme Court gets too mixed up in politics.
- The US Supreme Court should have the right to say what the Constitution means, even when the majority of the people disagree with the Court’s decisions.

B.4 Respondent Characteristics

Table SI.1: Respondent Descriptive Characteristics

<u>Characteristic</u>	Study 1 (MTurk)	Study 2 (Lucid)	Study 3 (Lucid)
<u>Age</u>			
18-29	23.3% (326)	24.5% (406)	26.7% (227)
30-49	54.6% (765)	38.4% (636)	35.7% (303)
50-64	16.6% (232)	21.5% (356)	22.5% (191)
65 and over	5.6% (78)	15.5% (257)	15.1% (128)
NA	0.0% (0)	0.0% (0)	0.0% (0)
<u>Gender</u>			
Female	53.7% (752)	51.2% (847)	51.2% (435)
Male	46.7% (649)	48.8% (808)	48.8% (414)
Other	0.0% (0)	0.0% (0)	0.0% (0)
NA	0.0% (0)	0.0% (0)	0.0% (0)
<u>Race/Ethnicity</u>			
White/Non-Hispanic	72.9% (1021)	66.5% (1100)	64.1% (544)
White/Hispanic	5.2% (73)	4.3% (71)	6.6% (56)
Black/Non-Hispanic	9.2% (129)	10.9% (181)	11.8% (100)
Black/Hispanic	2.5% (35)	2.1% (34)	2.5% (21)
Asian	6.7% (94)	4.4% (72)	3.4% (29)
Other	3.4% (48)	11.7% (193)	11.2% (95)
NA	0.1% (1)	0.2% (4)	0.5% (4)
<u>Education</u>			
High school degree or less	10.7% (150)	24.5% (406)	26.1 (221)%
Some college, no 4-year degree	31.8% (446)	29.8% (493)	30.4% (258)
Bachelor's degree	41.5% (581)	21.4% (354)	19.3 (164)%
Post-graduate degree	16.0% (224)	22.9% (379)	23.3% (198)
NA	0.0% (N)	1.4% (23)	0.9% (8)
<u>Income</u>			
Less than \$25,000	13.3% (187)	26.3% (435)	29.8% (253)
\$25,000-\$50,000	29.3% (410)	21.6% (357)	18.0% (153)
\$50,000-\$75,000	24.0% (336)	14.2% (235)	12.4 (105)%
\$75,000-\$100,000	17.0% (238)	8.3% (138)	10.0 (85)%
\$100,000-\$200,000	14.3% (200)	17.6% (292)	16.1 (137)%
\$200,000 or more	2.1% (29)	6.9% (114)	6.6 (56)%
NA	0.1% (1)	5.1% (84)	7.1 (60)%
<u>Party Identification</u>			

<u>Characteristic</u>	Nomination (MTurk)	Lower Court (Lucid)	Opinion (Lucid)
Democrat	53.5% (747)	41.0% (679)	43.7% (371)
Independent	10.9% (153)	8.5% (140)	8.8% (75)
Republican	33.0% (463)	44.9% (743)	42.2% (358)
Other	2.7% (38)	5.6% (93)	5.3% (45)
NA	0.0% (0)	0.0% (0)	0.0%
<u>Ideology</u>			
Very liberal	11.1% (155)	17.7% (293)	17.0% (144)
Liberal	22.6% (317)	14.7% (243)	14.5% (123)
Slightly liberal	12.5% (175)	8.2% (135)	8.7% (74)
Moderate	20.1% (281)	33.2% (549)	31.1% (264)
Slightly conservative	10.8% (152)	6.8% (113)	6.7% (57)
Conservative	16.1% (226)	11.1% (184)	12.8% (109)
Very conservative	6.8% (95)	8.2% (136)	8.6% (73)
NA	0.0% (N)	0.1% (2)	0.6% (5)

Note: This table indicates the percentage and number of respondents in each sample (denoted by the column headings) who reported each demographic characteristic (denoted by the row labels). Some characteristics may not sum to 100% due to rounding.

C Empirical Analyses

In this section, we describe and present our empirical analyses and discuss the interpretation of our null findings. The analysis we present in Figure 1 is based on the models summarized in Table SI.2. For specific support, we dichotomized our outcomes by coding respondents as approving (1) of the judge or case featured in the vignette if they “strongly approve” or “somewhat approve,” and not approving (0) otherwise. For diffuse support, we follow Gibson, Caldeira, and Spence (2003) and construct a scale using the six statements they introduced which ask respondents about their level of agreement with contours of the design of the Supreme Court/federal courts. Specifically, for each statement, we code respondents as supporting the judiciary (1) if they “strongly agree” or “somewhat agree” with a statement that expresses loyalty to the judiciary or “strongly disagree” or “somewhat disagree” with a statement that expresses disloyalty to the judiciary, and not supporting the judiciary (0) otherwise. Respondents’ reactions to the six statements are then summed and rescaled from 0-1. The Cronbach’s α for the diffuse support scales in our Supreme Court nomination, lower court, and Supreme Court opinion experiments are 0.69, 0.60, and 0.63, respectively.³⁸

³⁸ These moderately strong values are similar to those for scales used in recently published work on diffuse support (Gibson and Caldeira 2009*b*; Gibson and Nelson 2017, 2018).

All analyses include all respondents irrespective of attention check passage. The substantive interpretation of our findings is consistent across all three experiments when we use information about attention check passage to calculate complier average causal effects.³⁹

C.1 Robustness Checks

We provide additional analyses to demonstrate the robustness of our results to alternative specifications. The substantive interpretation of our findings is consistent across each of these specifications:

- In Table SI.4, we estimate the effects of our treatments on specific support in all three experiments using the original four-point ordinal scale as our outcome measure.
- In Tables SI.5 and SI.6, we estimate the effects of our treatments on specific support in our lower court and Supreme Court opinion experiments using multinomial logistic regression, which allows us to incorporate “don’t know” responses in our analysis.

C.2 Additional Randomizations

We also provide additional analyses that include the other randomized facets of our experiments—the partisan affiliation of the president who appointed the judge in our Supreme Court nomination and lower court experiments and the size of the Court’s majority in the Supreme Court opinion experiment. In part, these models allow us to investigate whether our results are driven by correspondence of partisan affiliation between the respondent and the featured judge (i.e., partisan motivated reasoning). The substantive interpretation of our findings is consistent across each of these alternative specifications and we do not find evidence that respondent copartisanship or majority coalition size condition the effect of scandal on specific or diffuse support:

- In Table SI.7, we estimate the effects of all of our randomized facets on specific and diffuse support.
- In Table SI.8, we estimate the copartisan-conditional effects of our scandal treatments by interacting respondents’ treatment status with a dichotomous indicator for whether they share the partisan affiliation of the president who appointed the featured judge.

C.3 Partisanship-Conditional Effects

We also conduct additional analyses that interact respondents’ partisan identification with our scandal treatments to probe whether certain types of scandals have heterogeneous effects

³⁹ Specifically, we re-estimate the models in Table SI.2, which underlie Figure 1, to obtain the complier average causal effects (CACEs) by using an instrumental variables framework with treatment assignment as our instrument, attentiveness to the treatment as our treatment variable (i.e., whether the respondent was provided with the treatment vignette *and* passed both attention checks), and specific/diffuse support as our outcome measures. These analyses are not presented here, but the code to replicate them is included in our replication materials. While the substantive magnitudes of the CACEs are naturally larger than the treatment effects presented in Figure 1 which do not account for respondents’ attentiveness, the statistical significance of the corresponding estimates are consistent with those presented in Table SI.2.

across respondents with different partisan affiliations. Given that Democrats and Republicans often express different beliefs and values concerning moral behavior and how they expect their elected officials to behave, their reactions to our scandal treatments might differ. For instance, several recent studies indicate that Democrats are more likely to agree with the tenets of the #MeToo movement and to punish political figures more strongly for sexual misconduct (Costa et al. 2020; Holman and Kalmoe 2021). These alternative specifications, presented in Table SI.9, indicate that partisan differences consistently manifest across Studies 1 and 2 only for the effect of sexual scandals on specific support, with Democratic respondents punishing judges accused of sexual misconduct more harshly than Republicans. However, we observe no consistent partisanship-conditional effects across scandal types for diffuse support.

C.4 Interpreting Null Results

While our experiments detect negative effects of scandal on specific support for scandalized judges, they yield null effects for the majority of our outcomes. Interpreting null effects can be difficult because they do not provide evidence that the treatment has no effect, but rather indicate that we cannot reject the null hypothesis that the treatment has no effect. Thus, researchers confronted with null effects must consider whether these nulls arise as type II errors or as manifestations of true null effects. While we cannot definitively demonstrate that our null effects represent true nulls, we discuss here that our null results pertaining to specific support in Study 3 and diffuse support in all three studies are unlikely to have arisen from design choices.

Specific Support in Study 3

In Study 3, we find a null effect for respondents' specific support for the Court's decision in *Ramos v. Louisiana* when Gorsuch or Kavanaugh is attributed as the opinion author. This specific support effect differs from those in Studies 1 and 2, but the designs of these studies differ from Study 3 in purposeful ways which may underlie the differences in findings. For instance, the objects for which specific support is expressed—an opinion authored by a specific judge rather than a specific judge—are different across studies, and it is possible that the mere presence of a scandalized judge as part of a panel rendering a decision does not exert negative effects on public opinion but the presence of the scandalized judge does. Second, whereas Studies 1 and 2 provide vignettes which explicitly mention the scandal experienced by the judge, Study 3 leverages the real-world political milieu to assess whether an actor implicated in a high-profile scandal taints judicial decisions with which he is associated in the future. Thus, the difference in specific support results on their own is not cause for concern that the specific support result in Study 3 is a false negative.

Another potential explanation for this finding being a type II error is that the treatment—assigning the opinion to Kavanaugh rather than Gorsuch—did not activate considerations related to the sexual assault allegations made against Kavanaugh, and thus judicial scandal did not influence respondents' evaluations. While we agree that our treatment may

not have made salient judicial scandal considerations, we argue that allowing for this possibility is an intended feature of our design rather than a bug that may lead to a type II error. As we elaborate in Section B.3.1, Study 3 intends to test for long-term effects of judicial scandal on public opinion in a real-world setting. These long-term effects are the ultimate concern of those expressing concern for the judiciary’s public support; while immediate public backlash against a judge involved in a scandal may be informative, the true threat to the judiciary’s legitimacy is if scandal can diminish support for its judges, rulings, and the institution as a whole once the scandal has passed. Thus, if respondents failed to call upon considerations concerning Kavanaugh’s sexual assault allegations when he is mentioned in a news story, this would suggest that, when it comes to the judiciary, penalties arising from scandal are short-lived and quickly pass out public memory.

The data available to us does not enable us to disentangle whether the Kavanaugh treatment may have yielded null effects because respondents did not consider the sexual assault allegations made against him or because they did consider those allegations but did not use them to inform their evaluations of the Court decision. While either mechanism would conform with our findings on the whole—that judicial scandal diminishes short-term specific support for judges but does not affect enduring evaluations of diffuse support—determining which of these explanations is at play is worthy of further research. We offer a rough test in Table SI.10 by interacting pre-treatment measures of respondents’ judicial knowledge with their treatment assignment; here, we expect that respondents with higher levels of judicial knowledge are more likely to have received information about the scandal surrounding Kavanaugh and thus more likely to call on those considerations when participating in our experiment.⁴⁰ However, the estimated conditional treatment effects (i.e., the Kavanaugh:Judicial Knowledge interaction term) are substantively small and statistically indistinguishable from 0, suggesting that even respondents more attuned to the Court did not respond negatively to Kavanaugh’s authorship of the opinion.

A final potential explanation for our null result for specific support in Study 3 is that respondents’ reactions to Justice Kavanaugh might be highly polarized due to his contentious confirmation hearings. While Democrats responded to the sexual assault allegations made against Kavanaugh with extreme negativity, many Republicans responded to the same allegations with fierce support in the belief that Kavanaugh was being unfairly maligned by political opponents with spurious allegations. Given the potential for strong polarization concerning Kavanaugh, copartisanship-conditional effects whereby Republican respondents responding positively to the mention of Kavanaugh and Democratic respondents responding negatively to Kavanaugh could offset yielding a null effect for the full sample. This is an important potential explanation to consider because, if supported by the data, it would suggest that judicial scandal might not diminish specific support for the full population, but can diminish specific support among respondents not aligned with the party associated with the implicated judge (i.e., the nominating president’s party) if the scandal was highly salient and unfolded in a highly partisan conflict.

⁴⁰ Judicial knowledge is measured by scaling respondents’ answers to the standard three-question battery used by the American National Election Studies (see Gibson and Caldeira 2009).

To assess this potential explanation, we turn to Table SI.8, which interacts each treatment with a dichotomous indicator for whether the respondent is a copartisan of the implicated judge’s nominating president. For Study 3, this means that Republicans are coded as copartisans (1) and Democrats, independents, and those identifying with any other party are coded as non-copartisans (0). If Kavanaugh engenders offsetting heterogeneous effects for specific support as described above, we would expect the coefficients on the “Kavanaugh:Copartisan” interaction term to be positive and statistically distinguishable, such that Republicans express higher levels of specific support than other respondents. However, the coefficients is substantively small and not statistically distinguishable, suggesting that our null finding is not a result of offsetting heterogeneous effects.

Diffuse Support in All Three Studies

Across our seven outcome measures of diffuse support from our three studies, only one scandal treatment (the effect of sexual scandal in Study 2) exerts a statistically distinguishable but substantively small effect on diffuse support. One potential concern about these null effects could be whether our design is sufficiently powered (i.e., are our null effects a type II error resulting from insufficient power?). Power calculations using the realized number of respondents in each treatment group indicate that the treatments in each of our studies are sufficiently powered to detect treatment effects of between 0.19 and 0.21 units on the 0-1 diffuse support scale.⁴¹ Given that this scale is constructed by summing binarized responses to six questions, the treatment effects we are sufficiently powered to detect are slightly larger than the effect of a respondent changing their answer on one of the six diffuse support questions (i.e., a one question change corresponds with a $0.1\bar{6}$ shift). We argue that having sufficient power to detect a one-question shift in diffuse support enables us to identify substantively important shifts in diffuse support caused by judicial scandal; while smaller treatment effects may exist, the magnitude of the shifts on the 0-1 scale they would represent would not indicate consequential effects of scandal on diffuse support.

However, we can also try improve the precision of our treatment effect estimates by including respondents’ pre-treatment covariates in our regression models (Gerber and Green 2012). In Table SI.3, we re-estimate the models we use to generate the treatment effect estimates in Figure 1 in the main paper by including a battery of pre-treatment covariates.⁴² The results are substantively similar when including covariates, further suggesting that any negative effects of judicial scandal on diffuse support, if they exist, are of small, inconsequential magnitude.

⁴¹ We conducted our power calculations using the `pwr` package in R using the true number of respondents in each treatment condition as our ns , $\alpha=0.05$, and $\beta=0.80$.

⁴² Our covariates include gender, education, ethnicity/race, income, party identification, ideology, and judicial knowledge.

Table SI.2: Effect of Judicial Scandal on Specific and Diffuse Support (OLS, 0-1 Scales)

	Specific Support			Diffuse Support		
	Nomination	Lower Court	Opinion	Nomination	Lower Court	Opinion
Intercept	0.85*	0.69*	0.58*	0.63*	0.47*	0.46*
	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Ethics Scandal	-0.35*	-0.29*		-0.04	-0.02	
	(0.04)	(0.03)		(0.02)	(0.02)	
Sexual Scandal	-0.46*	-0.24*		-0.02	-0.05*	
	(0.04)	(0.03)		(0.02)	(0.02)	
Tax Scandal	-0.24*	-0.25*		-0.02	-0.02	
	(0.04)	(0.03)		(0.02)	(0.02)	
Kavanaugh			0.06			0.03
			(0.03)			(0.02)
Num. Obs.	1399	1628	848	1401	1655	849

*Denotes statistical significance at the $p < 0.05$ level. Models estimated using ordinary least squares regression (OLS). Models include all respondents, irrespective of attention check passage. Specific support outcomes are measured with binary indicators where responses indicating that respondents somewhat or strongly approve are coded as 1 and all other non-missing responses coded as 0. Diffuse support is coded following Gibson, Caldeira, and Spence (2003) where we dichotomize respondents' answers to indicate support for the judiciary or lack thereof, sum the binary indicators, and rescale the final measure to range between 0 and 1.

Table SI.3: Effect of Judicial Scandal on Specific and Diffuse Support (OLS w/Covariates, 0-1 Scales)

	Specific Support			Diffuse Support		
	Nomination	Lower Court	Opinion	Nomination	Lower Court	Opinion
Intercept	0.83*	0.55*	0.26*	0.21*	0.27*	0.19*
	(0.08)	(0.07)	(0.09)	(0.05)	(0.04)	(0.05)
Ethics Scandal	-0.37*	-0.31*		-0.04*	-0.02	
	(0.03)	(0.03)		(0.02)	(0.02)	
Sexual Scandal	-0.46*	-0.26*		-0.03	-0.04*	
	(0.03)	(0.03)		(0.02)	(0.02)	
Financial Scandal	-0.25*	-0.26*		-0.03	-0.02	
	(0.03)	(0.03)		(0.02)	(0.02)	
Kavanaugh			0.05			0.02
			(0.03)			(0.02)
Num. obs.	1398	1627	843	1400	1653	844

*Denotes statistical significance at the $p < 0.05$ level. Models estimated using ordinary least squares regression (OLS) and include the following pretreatment covariates: gender, education, ethnicity/race, income, party identification, ideology, and judicial knowledge. Models include all respondents, irrespective of attention check passage. Specific support outcomes are measured with binary indicators where responses indicating that respondents somewhat or strongly approve are coded as 1 and all other non-missing responses coded as 0. Diffuse support is coded following Gibson, Caldeira, and Spence (2003) where we dichotomize respondents' answers to indicate support for the judiciary or lack thereof, sum the binary indicators, and rescale the final measure to range between 0 and 1.

Table SI.4: Effect of Judicial Scandal on Specific Support (OLS, Ordinal Scales)

	SCOTUS Nomination	Lower Court	SCOTUS Opinion
Intercept	3.09*	3.17*	2.93*
	(0.05)	(0.05)	(0.05)
Ethics Scandal	-0.62*	-0.72*	
	(0.06)	(0.07)	
Sexual Scandal	-0.87*	-0.61*	
	(0.06)	(0.07)	
Tax Scandal	-0.46*	-0.64*	
	(0.06)	(0.08)	
Kavanaugh			0.12
			(0.07)
Num. Obs.	1399	1366	698

*Denotes statistical significance at the $p < 0.05$ level. Models estimated using ordinary least squares regression (OLS). Models include all respondents, irrespective of attention check passage. Specific support outcomes are measured with ordinal indicators where responses on a 1-4 scale with respondents who strongly approve coded as 4 and respondents who strongly disapprove coded as 1. Respondents who indicated “Don’t Know” are coded as NAs and excluded from the analysis.

Table SI.5: Effect of Judicial Scandal on Specific Support of Lower Court Judge (Multinomial Logistic Regression)

	Strongly Disapp.	Somewhat Disapp.	Somewhat App.	Strongly App.
Intercept	-1.44*	-0.21	0.84*	0.81*
	(0.29)	(0.19)	(0.15)	(0.15)
Ethics Scandal	1.48*	0.52*	-0.53*	-0.92*
	(0.33)	(0.24)	(0.22)	(0.23)
Sexual Scandal	1.38*	0.46	-0.38	-0.77*
	(0.33)	(0.24)	(0.21)	(0.22)
Tax Scandal	1.56*	0.79*	-0.23	-0.59*
	(0.34)	(0.25)	(0.23)	(0.24)
Num. Obs.	1628			

*Denotes statistical significance at the $p < 0.05$ level. Models estimated using multinomial logistic regression to account for “Don’t know” responses in the two surveys fielded on Lucid. Models include all respondents, irrespective of attention check passage. Our outcome variable has five factor values — “Don’t Know” (the baseline choice), “Strongly Disapprove”, “Somewhat Disapprove”, “Somewhat Approve”, and “Strongly Approve”.

Table SI.6: Effect of Judicial Scandal on Specific Support of Supreme Court Opinion
(Multinomial Logistic Regression)

	Strongly Disapp.	Somewhat Disapp.	Somewhat App.	Strongly App.
Intercept	-1.09*	-0.15	0.56*	0.23
	(0.22)	(0.16)	(0.14)	(0.15)
Kavanaugh	-0.07	0.06	0.18	0.37
	(0.33)	(0.24)	(0.20)	(0.21)
Num. Obs.	848			

*Denotes statistical significance at the $p < 0.05$ level. Models estimated using multinomial logistic regression to account for “Don’t Know” responses in the two surveys fielded on Lucid. Models include all respondents, irrespective of attention check passage. Our outcome variable has five factor values — “Don’t Know” (the baseline), “Strongly Disapprove”, “Somewhat Disapprove”, “Somewhat Approve”, and “Strongly Approve”.

Table SI.7: Effect of Judicial Scandal on Specific and Diffuse Support (OLS, 0-1 Scales,
Including All Randomized Factors)

	Specific Support			Diffuse Support		
	Nomination	Lower Court	Opinion	Nomination	Lower Court	Opinion
Intercept	0.84*	0.66*	0.58*	0.65*	0.46*	0.46*
	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Ethics Scandal	-0.40*	-0.28*		-0.03	-0.03	
	(0.05)	(0.05)		(0.03)	(0.03)	
Sexual Scandal	-0.47*	-0.20*		-0.04	-0.06*	
	(0.05)	(0.05)		(0.03)	(0.03)	
Tax Scandal	-0.27*	-0.25*		-0.05	-0.01	
	(0.05)	(0.05)		(0.03)	(0.03)	
Obama	0.01	0.05		-0.02	0.01	
	(0.05)	(0.05)		(0.03)	(0.03)	
Ethics Scandal: Obama	0.10	-0.02		-0.01	0.01	
	(0.07)	(0.07)		(0.04)	(0.04)	
Sexual Scandal: Obama	0.02	-0.09		0.03	0.01	
	(0.07)	(0.07)		(0.04)	(0.04)	
Tax Scandal: Obama	0.07	-0.00		0.06	-0.03	
	(0.07)	(0.07)		(0.04)	(0.04)	
Kavanaugh			0.02			0.05
			(0.05)			(0.03)
Unanimous			0.00			-0.01
			(0.05)			(0.03)
Kavanaugh: Unanimous			0.09			-0.04
			(0.07)			(0.04)
Num. Obs.	1399	1628	848	1401	1655	849

*Denotes statistical significance at the $p < 0.05$ level. Models estimated using ordinary least squares regression (OLS). Models include all respondents, irrespective of attention check passage. Specific support outcomes are measured with binary indicators where responses indicating that respondents somewhat or strongly approve are coded as 1 and all other non-missing responses coded as 0. Diffuse support is coded following Gibson, Caldeira, and Spence (2003) where we dichotomize respondents’ answers to indicate support for the judiciary or lack thereof, sum the binary indicators, and rescale the final measure to range between 0 and 1. All models interact our scandal treatments with the other factors we randomized in the vignettes —the appointing president in our nomination and lower court experiments (George W. Bush or Barack Obama) and the margin by which the Supreme Court decided *Ramos v. Louisiana* (9-0 or 6-3).

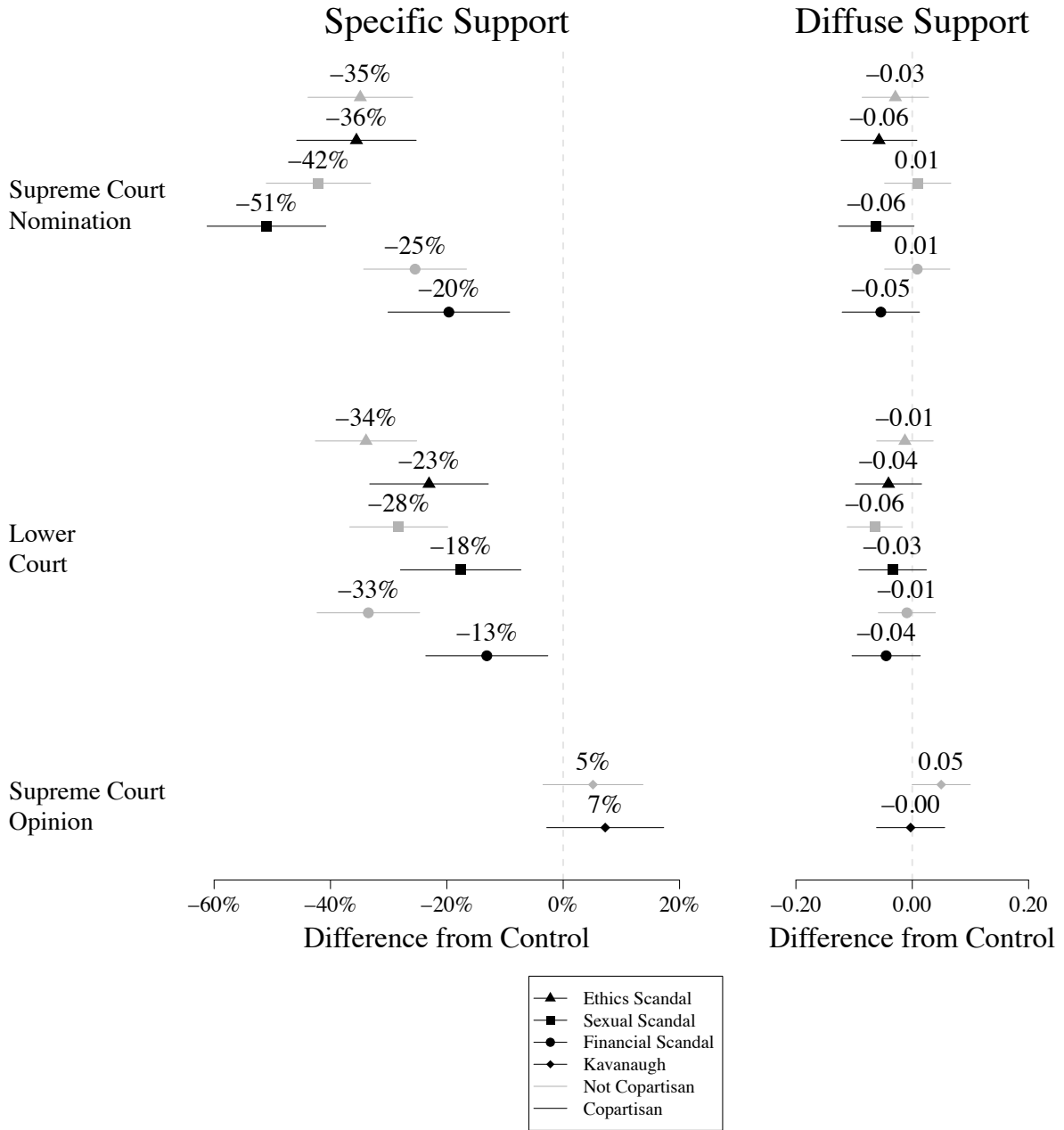


Figure SI.1: **Effect of Scandal on Specific and Diffuse Support (Copartisanship-Conditional)**. Linear regression coefficients for copartisanship-conditional treatments effects. Positive values along the x -axis reflect higher levels of support. Bars around point estimates represent 95% confidence intervals. Specific support is measured with a binary indicator for whether the respondents approved of the judge (Supreme Court Nomination and Lower Court) or opinion (Supreme Court opinion) featured in the vignette. Diffuse support is measured on a 0-1 scale using respondents' answers to the six questions used by (Gibson, Caldeira, and Spence 2003). Please see Supplemental Information B for vignette and question wordings. Please see Table SI.8 for regression summaries.

Table SI.8: Effect of Judicial Scandal on Specific and Diffuse Support (OLS, 0-1 Scales, Copartisanship-Conditional)

	Specific Support			Diffuse Support		
	Nomination	Lower Court	Opinion	Nomination	Lower Court	Opinion
Intercept	0.76*	0.69*	0.53*	0.61*	0.47*	0.45*
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Ethics Scandal	-0.35*	-0.34*		-0.03	-0.01	
	(0.05)	(0.04)		(0.03)	(0.02)	
Sexual Scandal	-0.42*	-0.28*		0.01	-0.06*	
	(0.05)	(0.04)		(0.03)	(0.02)	
Tax Scandal	-0.25*	-0.33*		0.01	-0.01	
	(0.05)	(0.04)		(0.03)	(0.03)	
Kavanaugh			0.05			0.05
			(0.04)			(0.03)
Copartisan	0.20*	-0.00	0.11*	0.06*	0.00	0.03
	(0.05)	(0.05)	(0.05)	(0.03)	(0.03)	(0.03)
Ethics Scandal: Copartisan	-0.01	0.11		-0.03	-0.03	
	(0.07)	(0.07)		(0.04)	(0.04)	
Sexual Scandal: Copartisan	-0.09	0.11		-0.07	0.03	
	(0.07)	(0.07)		(0.04)	(0.04)	
Tax Scandal: Copartisan	0.06	0.20*		-0.06	-0.04	
	(0.07)	(0.07)		(0.04)	(0.04)	
Kavanaugh: Copartisan			0.02			-0.05
			(0.07)			(0.04)
Num. Obs.	1399	1628	848	1401	1655	849

*Denotes statistical significance at the $p < 0.05$ level. Models estimated using ordinary least squares regression (OLS). Models include all respondents, irrespective of attention check passage. Specific support outcomes are measured with binary indicators where responses indicating that respondents somewhat or strongly approve are coded as 1 and all other non-missing responses coded as 0. Diffuse support is coded following Gibson, Caldeira, and Spence (2003) where we dichotomize respondents' answers to indicate support for the judiciary or lack thereof, sum the binary indicators, and rescale the final measure to range between 0 and 1. All models interact our scandal treatments with a binary indicator for whether the respondent shares the partisan affiliation of the featured judge. For the nomination and lower court experiments, respondents are coded as copartisans if they are Democrats and the president who appointed the judge was Barack Obama or if they are Republicans and the president who appointed the judge is George W. Bush. For the opinion experiment, respondents are coded as copartisans if they are Republicans (since both justices who could be featured were appointed by a Republican president).

Table SI.9: Effect of Judicial Scandal on Specific and Diffuse Support (OLS, 0-1 Scales, Partisanship-Conditional)

	Specific Support			Diffuse Support		
	Nomination	Lower Court	Opinion	Nomination	Lower Court	Opinion
Intercept	0.85*	0.70*	0.64*	0.68*	0.44*	0.47*
	(0.04)	(0.04)	(0.04)	(0.03)	(0.02)	(0.02)
Ethics Scandal	-0.27*	-0.21*		-0.05	0.00	
	(0.06)	(0.05)		(0.04)	(0.03)	
Sexual Scandal	-0.24*	-0.10		-0.04	-0.02	
	(0.06)	(0.05)		(0.04)	(0.03)	
Tax Scandal	-0.17*	-0.22*		0.01	0.02	
	(0.06)	(0.05)		(0.04)	(0.03)	
Kavanaugh			0.07			-0.00
			(0.05)			(0.03)
Democrat	0.02	0.02	-0.06	-0.06	0.09*	-0.01
	(0.05)	(0.05)	(0.05)	(0.03)	(0.03)	(0.03)
Independent	-0.05	0.01	-0.30*	-0.16*	0.02	-0.04
	(0.08)	(0.08)	(0.09)	(0.05)	(0.05)	(0.05)
Other	-0.07	-0.25*	-0.31*	0.06	-0.12*	-0.14*
	(0.16)	(0.10)	(0.11)	(0.10)	(0.06)	(0.07)
Ethics Scandal:	-0.14	-0.15*		-0.00	-0.07	
Democrat	(0.08)	(0.07)		(0.05)	(0.04)	
Sexual Scandal:	-0.34*	-0.27*		0.02	-0.07	
Democrat	(0.08)	(0.07)		(0.05)	(0.04)	
Tax Scandal:	-0.11	-0.06		-0.05	-0.09*	
Democrat	(0.08)	(0.07)		(0.05)	(0.04)	
Kavanaugh:			-0.01			0.06
Democrat			(0.07)			(0.04)
Ethics Scandal:	-0.07	-0.25		0.15	-0.08	
Independent	(0.12)	(0.13)		(0.08)	(0.07)	
Sexual Scandal:	-0.22	-0.30*		0.12	-0.02	
Independent	(0.12)	(0.12)		(0.08)	(0.06)	
Tax Scandal:	-0.05	-0.24		0.01	0.01	
Independent	(0.12)	(0.14)		(0.07)	(0.07)	
Kavanaugh:			0.00			0.04
Independent			(0.12)			(0.07)
Ethics Scandal:	-0.05	-0.15		-0.15	0.12	
Other	(0.21)	(0.15)		(0.13)	(0.08)	
Sexual Scandal:	0.09	-0.05		-0.06	0.02	
Other	(0.23)	(0.14)		(0.14)	(0.08)	
Tax Scandal:	-0.10	0.04		-0.23	-0.06	
Other	(0.23)	(0.15)		(0.14)	(0.08)	
Kavanaugh:			0.01			0.07
Other			(0.15)			(0.09)
Num. Obs.	1399	1628	848	1401	1655	849

*Denotes statistical significance at the $p < 0.05$ level. Models estimated using ordinary least squares regression (OLS). Models include all respondents, irrespective of attention check passage. Specific support outcomes are measured with binary indicators where responses indicating that respondents somewhat or strongly approve are coded as 1 and all other non-missing responses coded as 0. Diffuse support is coded following Gibson, Caldeira, and Spence (2003) where we dichotomize respondents' answers to indicate support for the judiciary or lack thereof, sum the binary indicators, and rescale the final measure to range between 0 and 1. All models interact our scandal treatments with binary indicators for whether the respondent identifies as a Democrat, Independent, or an unspecified party (Other), with identification as a Republican as the reference category.

Table SI.10: Effect of Kavanaugh Authorship on Specific and Diffuse Support (OLS, 0-1 Scales, Knowledge-Conditional)

	Specific Support	Diffuse Support
Intercept	0.45* (0.04)	0.30* (0.02)
Kavanaugh	0.05 (0.06)	-0.01 (0.03)
Judicial Knowledge	0.08* (0.02)	0.09* (0.01)
Kavanaugh:Judicial Knowledge	0.00 (0.03)	0.02 (0.02)
Num. Obs.	848	849

*Denotes statistical significance at the $p < 0.05$ level. Models estimated using ordinary least squares regression (OLS). Models include all respondents, irrespective of attention check passage. Specific support outcome is measured with a binary indicator where responses indicating that respondents somewhat or strongly approve are coded as 1 and all other non-missing responses coded as 0. Diffuse support is coded following Gibson, Caldeira, and Spence (2003) where we dichotomize respondents' answers to indicate support for the judiciary or lack thereof, sum the binary indicators, and rescale the final measure to range between 0 and 1. Judicial knowledge is measured by scaling respondents' answers to the standard three-question battery used by the American National Election Studies (see Gibson and Caldeira 2009).

References

- Bartels, Brandon, and Christopher Johnston. 2013. "On the Ideological Foundations of Supreme Court Legitimacy in the American Public." *American Journal of Political Science* 57(1): 184–199.
- Basinger, Scott, Lara Brown, Douglas Harris, and Girish Gulati. 2013. "Preface: Counting and Classifying Congressional Scandals." In *Scandal!: An Interdisciplinary Approach to the Consequences, Outcomes, and Significance of Political Scandals*, ed. Alison Dagnes, and Mark Sachleben. 7 ed. Bloomsbury.
- Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20(3): 351–368.
- Berinsky, Adam J., Michele F. Margolis, and Michael W. Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58(3): 739–753.
- Coppock, Alexander. 2019. "Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach." *Political Science Research and Methods* 7(3): 613–628.

- Coppock, Alexander, and Oliver A. McClellan. 2019. “Validating the Demographic, Political, Psychological, and Experimental Results Obtained from a New Source of Online Survey Respondents.” *Research & Politics* 6(1): 1–14.
- Coppock, Alexander, Thomas J. Leeper, and Kevin J. Mullinix. 2018. “Generalizability of Heterogeneous Treatment Effect Estimates Across Samples.” *Proceedings of the National Academy of Sciences* 115(49): 12441–12446.
- Costa, Mia, Trevor Briggs, Ajaipal Chahal, Jonathan Fried, Rijul Garg, Sophia Kriz, Leo Lei, Anthony Milne, and Jenna Slayton. 2020. “How Partisanship and Sexism Influence Voters’ Reactions to Political #MeToo Scandals.” *Research & Politics* 7(3): 1–8.
- Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. WW Norton.
- Gibson, James, and Gregory A. Caldeira. 2009a. *Citizens, Courts, and Confirmations: Positivity Theory and the Judgments of the American People*. Princeton University Press.
- Gibson, James, and Gregory Caldeira. 2009b. “Confirmation Politics and the Legitimacy of the US Supreme Court: Institutional Loyalty, Positivity Bias, and the Alito Nomination.” *American Journal of Political Science* 53(1): 139–155.
- Gibson, James, and Gregory Caldeira. 2013. “Judicial Impartiality, Campaign Contributions, and Recusals: Results from a National Survey.” *Journal of Empirical Legal Studies* 10(1): 76–103.
- Gibson, James, and Michael J. Nelson. 2017. “Reconsidering Positivity Theory: What Roles Do Politicization, Ideological Disagreement, and Legal Realism Play in Shaping US Supreme Court Legitimacy?” *Journal of Empirical Legal Studies* 14(3): 592–617.
- Gibson, James, and Michael J. Nelson. 2018. *Blacks and Blue: How and Why African Americans Judge the American Legal System*. Oxford University Press.
- Gibson, James, Gregory Caldeira, and Lester Kenyatta Spence. 2003. “Measuring Attitudes Toward the United States Supreme Court.” *American Journal of Political Science* 47(2): 354–367.
- Holman, Mirya R., and Nathan P. Kalmoe. 2021. “Partisanship in the #MeToo Era.” *Perspectives on Politics*. Published ahead of print July 29, 2021.
- Kennedy, Ryan, Scott Clifford, Tyler Burleigh, Philip D. Waggoner, Ryan Jewell, and Nicholas J. G. Winter. 2020. “The Shape of and Solutions to the MTurk Quality Crisis.” *Political Science Research and Methods* 8(4): 614–629.
- Nelson, Michael. 2018. “Is There A Silver Lining? Dark Money and Support for State Courts.” *DePaul Law Review* 67(2): 187–218.

Ono, Yoshikuni, and Michael Zilis. 2022. "Ascriptive Characteristics and Perceptions of Impropriety in the Rule of Law: Race, Gender, and Public Assessments of Whether Judges Can Be Impartial." *American Journal of Political Science* 66(1): 43–58.