On the misplaced politics of behavioral policy interventions*

David Tannenbaum University of Chicago Craig R. Fox UCLA

Todd Rogers Harvard University

Abstract

One common criticism of policy tools derived from behavioral science that can be applied to a range of policy objectives (commonly referred to as *nudges*) is that such interventions are manipulative and coercive. We show that such criticism sometimes reflects a *partisan nudge bias*, whereby partisans conflate their attitudes toward policy tools with their attitudes toward policy objectives. Across five experiments we find that partisans evaluate policy nudges as relatively unethical when applied to policy objectives they oppose or by policymakers they distrust, but evaluate the same policy tools more favorably when applied to political objectives they support or by policymakers they trust. Both politically liberal and conservative partisans exhibited partisan nudge bias, as did practicing policymakers. However, these partisan differences disappear when behavioral policy interventions are described without a specific policy application or sponsor. Thus, explicitly setting aside particular policy objectives and endorsements may facilitate less prejudiced discussion about the appropriateness of deploying behavioral policy tools.

^{*}We are grateful to David King at the Harvard Kennedy School for assistance in data collection for Study 3A, and to Catherine McLaughlin and Christian Flynn at the Harvard Institute of Politics for their assistance in data collection on Study 3B. We also thank Dan Walters and Carsten Erner for their useful comments on an earlier draft of this paper. Correspondence should be addressed to David Tannenbaum, Chicago Booth School of Business, 5807 S. Woodlawn Ave, Rm 373. Email: david.tannenbaum@chicagobooth.edu.

Introduction

Insights from the behavioral and social sciences have recently been put to use in crafting effective public policy. These approaches are typically designed to "nudge" behavior that policymakers believe will promote individual or societal interests while also preserving individual freedom to choose.¹ For instance, one powerful behavioral insight is the tendency for people to stay with options designated as the default. As such, policies promoting automatic enrollment of employees into retirement saving plans with the ability to opt-out of the plan (rather than enrolling them only if they opt-in) have substantially increased 401(k) retirement saving rates in the United States.^{2,3} Such nudges are often cost-effective relative to other policy approaches (like education campaigns) and impose minimal costs on individuals with well-established preferences.⁴ Nudges do all this without the heavy-handedness and complexity of policy mandates or outright sanctions.

Despite these virtues, the use of policy nudges has sparked controversy, with opponents claiming that such techniques are manipulative and coercive. In the United Kingdom, where the conservative administration of David Cameron had been at the forefront in applying behavioral insights to public policy, criticism had been raised mostly by the political left. For instance, one contributor for the left-of-center publication *The Guardian* remarked that "however sympathetic we are to the goals a nudge is trying to achieve... we should be deeply skeptical of its tactics, which involve influencing the public without them knowing it is happening." Meanwhile, when the Obama administration announced the creation of a U.S. Social and Behavioral Sciences Team it also raised concerns of manipulation and coercion, this time mostly from the political right. For instance, Fox News contributor Monica Crowley described the initiative as "an Orwellian horror show," while Lou Dobbs remarked that such efforts sound "purely like propaganda and mind control."

As these examples suggest, objections to the application of behavioral policy interventions do not appear to fall along consistent ideological lines. In the U.K., advocates of smaller government have embraced the use of behavioral science as a means of enhancing policy efficiency by empowering individual choice; in the U.S., many individuals who advocate for smaller government have viewed such applications as threats to individual autonomy. This disparity of views raises the question of what, exactly, is it that people find ethically problematic about nudges? Criticisms of nudges by opposing political parties in the U.K. and U.S. suggests a reflexive and partisan rather than reflective and principled source.

In this article we examine whether laypeople and policymakers exhibit a *partisan nudge bias* in which attitudes toward behavioral policy tools (i.e., the means) are conflated with attitudes toward the particular policy to which it is applied (i.e., the ends) or politicians endorsing that policy. To illustrate this distinction, it is useful to consider instances where there is widespread support for a policy objective (for example, encouraging legal voter turnout) yet also strong agreement that some means for furthering that objective are ethically inappropriate (for example, threatening citizens with physical violence if they fail to vote). While attitudes toward policy means are clearly separable from attitudes toward policy ends, we examine whether people unwittingly confuse the two when

applied to politically-charged policy objectives. In particular, we provided both laypeople and current policymakers with short descriptions of nudges and randomly varied whether the nudge was applied to a particular policy application or endorsed by a particular political administration. We then asked them to evaluate whether such tools were manipulative, coercive, and unethical (see Table 1 for an overview of our studies). Partisans were considered biased if they viewed the use of a nudge (as a general-purpose policy tool) less favorably when applied to policy objectives they opposed rather than supported (or by policymakers they opposed rather than supported).

The studies described in this paper rely on brief sketches of policies because doing so provides the same level of detail, if not greater, presented in typical newspaper articles describing policymaking and which provide fodder for most political discourse. Moreover, numerous research studies have shown that even thoughtful and well-considered judgments can be profoundly shaped by first impressions. Thus, examining initial reactions to brief policy descriptions provides a credible indication of how the public and policymakers would evaluate policy nudges if given extended opportunities to familiarize themselves with greater detail.

Our research question is not merely an academic exercise. If resistance to the use of policy nudges is consistent and principled then we can and should focus debate on the acceptability of their use by policymakers and practitioners. But if resistance to policy nudges reflects misplaced discomfort with the particular policy application or advocate, then debate should be refocused on the relevant policy objectives. Demonstrating that nudges are not inherently partisan tools should make them more acceptable to a broad range of governing coalitions.

Study 1: Evaluations of Nudges are Influenced by the Policymaker Endorsing Them

Do partisans selectively disapprove of policy nudges when they are implemented by policymakers they distrust? To answer this question, we asked participants to evaluate a key provision of the 2006 Pension Protection Act (PPA) that encouraged employee retirement savings behavior through the use of automatic enrollment defaults. The PPA was endorsed and implemented by both the George W. Bush and Barack Obama administrations, providing a straightforward test case for examining how information about specific policymakers influences attitudes toward a popular policy nudge (i.e., the adoption of automatic enrollment defaults). We simply mentioned to participants that: (a) the Bush administration had enforced the law, (b) the Obama administration had enforced the law, or (c) neither (we omitted information about policymakers altogether as a control condition). If individuals confuse their attitudes toward the policymaker with their attitudes toward the policy nudge, then liberals should view automatic enrollment policies as unethical when implemented by the Bush administration but as ethical when implemented by the Obama administration. Conversely, conservatives should show the reverse pattern.

Methods Our sample consisted of 355 adults who were unfamiliar with the Pension Protection Act (54% male, mean age = 30 years, range: 18–67 years), using Amazon.com's Mechanical Turk online labor pool. ^{9,10,11} Compared to the general U.S. population, participants in this labor pool

Table 1: Overview of Studies

	Research Question	Population Sample	Policy Nudge	Experimental Treatment	Summary of Results
Study 1	Policymaker Bias: Do people conflate their attitudes towards a policy nudge with their attitudes towards a policymaker?	Online adults from Mechanical Turk labor market	Automatic enrollment provision of the 2006 Public Protection Act	1) Obama implements PPA 2) Bush implements PPA 3) No policymaker info provided	Liberals tended to favor auto-enrollment as a general policy tool when implemented under Obama compared to Bush; Conservatives showed the opposite tendency
Study 2	Policy Application Bias: Do people conflate their attitudes towards a policy nudge with their attitudes towards policy objectives?	Online adults from Mechanical Turk labor market	1) Automatic enrollment defaults 2) Planning prompts 3) Leveraging loss aversion 4) Prompt public commitments 5) Descriptive social norms	1) control/no example provided 2) food stamps 3) tax breaks 4) safe sex education 5) intelligent design education	Liberals tended to favor policy nudges as general policy tools when illustrated with applications to liberal policy objectives compared to conservative policy objectives; Conservatives showed the opposite tendency
Study 3A	Do experienced policymakers show a partisan nudge bias?	High-level bureaucratic leaders	Automatic enrollment defaults	1) low-income citizens automatically receive food stamps but can opt-out 2) high-income citizens automatically receive tax breaks but can opt-out	Liberals tended to favor auto-enrollment as a general policy tool when illustrated by applications to supplemental nutrition assistance for the poor compared to applications to tax breaks for the wealthy; Conservatives showed the opposite tendency
Study 3B	Do practicing policymakers show a partisan nudge bias?	U.S. Mayors	Automatic enrollment defaults	1) high school students automatically receive safe sex education, but parents can opt-out 2) high school students automatically receive education on intelligent design, but parents can opt-out	Liberals tended to favor auto-enrollment as a general policy tool when illustrated by applications to increasing safe sex education compared to applications to intelligent design education; Conservatives showed the opposite tendency

tend to skew female, are younger and better educated, and earn lower incomes.¹²

All participants read a brief statement in plain, neutral language about how the strategic selection of default options could be used as a general policy tool. As an illustration of the concept, participants were told that the U.S. Congress had recently passed a law (the PPA) that encouraged companies to automatically enroll employees into retirement savings plans while allowing employees to opt-out of the plan if they wished to do so. We randomly assigned participants to illustrations mentioning that the PPA had been enforced under the George W. Bush administration, the Barack Obama administration, or that it had simply been enforced by "lawmakers."

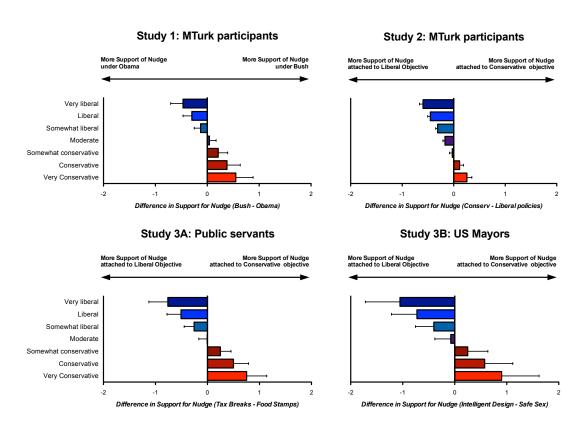
After reading about the policy nudge, we reminded participants that the strategic selection of defaults could be used "across a wide range of policies beyond the illustration above" and asked how they felt, setting the particular application aside, "about actively setting default options as a general approach to public policy" [italics in original text]. Participants then responded to six questions asking how much they supported and opposed using defaults as a general approach to public policy, and how ethical, unethical, coercive, and manipulative they found their use to be (e.g., 1 = strongly oppose, 5 = strongly support). All items were combined and averaged to form a highly reliable index of nudge evaluations (Cronbach's α = .89), with higher numbers indicating greater endorsement.

At the end of the study, participants rated their political orientation both for economic and social issues ($1 = very \ liberal$, $7 = very \ conservative$) and we averaged the two responses to create a single index of political orientation. Finally, we asked participants to separately rate the degree to which they found George W. Bush and Barack Obama trustworthy ($1 = not \ at \ all \ trustworthy$). 13

Results As we predicted, liberal and conservative participants evaluated the use of automatic enrollment defaults differently depending on which president was applying them, thereby confirming the existence of a partisan nudge bias. When participants were informed that the PPA had been enforced under the Obama administration, conservatives were more likely than liberals to oppose the use of defaults as a general policy tool ($\beta = -.21$, p = .017). However, this pattern was eliminated when participants were told that the same policy nudge had been enforced under the Bush administration ($\beta = .05$, p = .60), and the difference across the two conditions was statistically reliable (p = .05 when comparing the two slopes).

The upper-left panel of Figure 1 displays the categories of political orientation from very liberal (dark blue) at the top to very conservative (bright red) at the bottom. Each bar represents the difference in evaluations of automatic enrollment when the PPA was described as an Obama versus Bush program (values are smoothed by deriving the linear trend that best fits the data). Negative values imply a partisan nudge bias in favor of "Obama," whereas positive values imply a partisan nudge bias in favor of "Bush." The Figure shows that differences in evaluations are largest among respondents with the most extreme political views (bars at the top and bottom), and virtually absent among respondents who described themselves as moderate (bars closer to the middle).

Figure 1: Expected differences in evaluation of nudges. Values are smoothed by deriving the linear trend that best fits the data. For Study 1, difference scores represent evaluations under Bush vs. Obama. For Study 2, difference scores represent evaluations when illustrated by liberal versus conservative policy objectives (with evaluations aggregated across all policy nudges). For Study 3A, difference scores represent evaluations when illustrated by tax breaks vs. food stamps. For Study 3B, difference scores represent evaluations when illustrated by intelligent design vs. safe sex. Error bars represent robust standard errors.



Turning to the control condition in which information about the particular endorsing administration was omitted, we observed no apparent relationship between political orientation and nudge evaluations ($\beta = -.06$, p = .56). Thus, liberals and conservatives did not systematically disagree about the appropriateness of using defaults as a general policy tool when they were unaware of which political administration was applying the nudge.

The foregoing results suggest that partisan nudge bias may be driven by respondents' general feelings of trust toward the agent applying the nudge. If so, our direct measures of trust towards each policymaker should provide even stronger results than our indirect measure of political orientation. Indeed, ratings of trust in Obama were positively associated with nudge evaluations when participants had been informed that Obama enforced the PPA ($\beta = .47$, p < .001), but not when they were informed that Bush had enforced the PPA ($\beta = .00$, p = .97). Similarly, ratings of trust in Bush were positively associated with nudge evaluations when participants were informed that Bush had enforced the PPA ($\beta = .24$, p = .008), but not when they were informed that Obama had enforced the PPA ($\beta = -.10$, p = .30).

Study 2: Evaluations of Nudges are Influenced by the Policies Used to Illustrate Them

Study 1 suggests that people sometimes conflate their attitudes toward policy tools with their attitudes toward the policymakers using those tools. We next investigate whether this pattern extends to differences in attitudes toward the political objectives being advanced by the nudge. If people have difficulty separating their attitudes toward policy means from their attitudes toward policy ends, then we would expect participants to find policy nudges more unpalatable when illustrated with policy objectives they oppose than policy objectives they support.

In addition to automatic enrollment defaults, we examined four empirically established, field-tested policy nudges: (1) *implementation intentions*: prompting individuals to articulate concrete action plans to enhance follow-through, ^{15,16} (2) *highlighting losses*: strategically underscoring the potential costs of undesired behavior rather than the potential benefits of desired behavior, ¹⁷ leveraging the well-established tendency for losses to loom psychologically larger than equivalent gains, ¹⁸ (3) *commitment and consistency*: prompting individuals to publicly commit to behaviors in advance, ^{19,20} and (4) *descriptive social norms*: providing individuals with information about how other people behave and decide. ^{21,22} By experimentally varying the policy objective attached to each nudge, we can identify the independent effect each political objective has on evaluations of policy nudges for liberals and conservatives.

A secondary goal of Study 2 was to compare the relative impact of a partisan nudge bias to more established and principled attitudes concerning the role of government on individual choice. In particular, we asked individuals to rate their endorsement of libertarian versus paternalistic values. We expected that libertarian-leaning respondents would be more likely to oppose the use of nudges than respondents who endorsed more paternalistic values. Measuring libertarian/paternalistic values provides a rough benchmark of the "principled" effect of values on support for various nudges, against which we can evaluate the magnitude of any observed partisan nudge bias.

Methods We recruited a new sample of 462 adults from the same subject pool used in Study 1 (67% male, mean age = 33 years, range: 19–74 years). Participants responded to five different types of policy nudges: (1) automatic enrollment defaults, (2) implementation intentions, (3) public commitments, (4) highlighting losses, and (5) descriptive social norms. All study materials are provided in the Supplemental Material.

Participants were provided with brief descriptions of the nature of each policy nudge, which were randomly paired with an example application illustrating the use of the nudge. Illustrative policy applications included: (a) increasing participation by low-income individuals in existing food stamp and supplemental nutrition assistance programs ("food stamps"), (b) increasing participation of claims by high-income individuals to existing tax breaks on earned capital gains ("tax breaks"), (c) increasing participation in safe sex and effective contraceptive-use educational programs for high-school students ("safe sex"), (d) increasing participation in intelligent design educational programs for high-school students ("intelligent design"), and (e) a generic, context-free policy illustration ("control illustration"). The first four policy applications were designed to appeal to

economic liberals, economic conservatives, social liberals, and social conservatives, respectively. Each policy description was presented on a separate page, and we randomized the order in which policy nudges were presented as well as the specific policy application used to illustrate the nudge. Participants viewed each nudge and policy illustration once and only once; thus, individual participants responded to 5 policies from a pool of 25 possible pairings.

After reading about each policy nudge, participants were reminded that the approach was general and could be used across a wide range of contexts. Next, they evaluated the ethical acceptability of each nudge using the same six questions from Study 1. Again, we combined these six questions to form a single index of nudge evaluations (Cronbach's α = .89 when pooling over all observations). Afterward, participants reported their political orientation in the same manner as in Study 1. We also asked them to complete a six-question scale that measured individual differences in Libertarian attitudes²³ (e.g., "It's not the government's business to try to protect people from themselves," "Sometimes government needs to make laws that keep people from hurting themselves"). Participants rated their agreement with each statement (1 = strongly disagree, 7 = strongly agree), and all responses were combined, with higher scores indicating greater endorsement of libertarianism (Cronbach's α = .83).

Results As predicted, support for the use of policy nudges depended on the policy goals used to illustrate them.²⁴ When nudges were applied to traditionally liberal policy goals (increasing participation in food stamps programs or enrollment in safe sex education initiatives), conservatives were more likely than liberals find them ethically problematic ($\beta = .20$, p < .001 for food stamps; $\beta = .08$, p = .071 for safe sex). In contrast, liberals were more likely than conservatives to find these same nudges ethically problematic when they were applied to traditionally conservative policy goals ($\beta = .09$, p = .045 for tax breaks; $\beta = .08$, p = .069 for intelligent design). The upper-right panel of Figure 1 illustrates this partisan nudge bias. Participants with more extreme political orientations (top and bottom bars) tended to be more inconsistent in their evaluation of nudges across different applications, while political moderates tended to be less affected by the application (bars closer to the middle).

Another way to look at the data are to compare each individual's attitudes toward nudges that were illustrated using the two traditionally liberal policies versus their attitudes toward nudges that were illustrated using the two traditionally conservative policies.²⁵ Using this alternative within-participant analysis we again find evidence of partisan nudge bias, with participants displaying greater overall support for nudges when illustrated using policy objectives aligned with their political leanings than when illustrated by objectives at odds with their political leanings ($\beta = .24$, p < .001).

Similar to our analysis of trust in Study 1, we observe stronger indication of nudge bias when using direct measures of attitudes toward specific policy objectives, rather than a general measure of overall political orientation. For every application, we found a significant positive relationship between attitudes toward the corresponding policy goal and support for using nudges as general

policy tools (β -weights ranged from .16 to .46; all p-values < .01).

Turning to the analysis of nudges when they were not illustrated with a policy objective (the control illustration), we once again find no partisan differences in evaluations ($\beta = .00$, p = .89). Just as in Study 1, liberals and conservatives apparently find nudges equally acceptable when described without a partisan policy illustration. However, we do find that individual difference in Libertarian values do, in fact, predict attitudes toward generic descriptions of nudges ($\beta = -.23$, p < .001).

Interestingly, looking across policies we find that partisan nudge bias is considerably stronger than the (presumably more principled) effects of Libertarian attitudes in explaining how participants rated acceptability of nudges. When we simultaneously examine the impact of each variable on nudge evaluations (using partial correlations), we find that that the latter was over three times stronger as that of Libertarian attitudes (r = .30 and -.09, respectively).

Study 3: Policymakers Exhibit Partisan Nudge Bias

Study 2 demonstrated partisan nudge bias among respondents from an online sample, but the question arises whether this bias also extends to respondents who actually shape and implement policy. Experienced policymakers should have well-formed and well-informed preferences concerning the use different policy tools, since they regularly participate in discussions about how to best accomplish policy objectives. To examine whether policymakers exhibit partisan nudge bias, we sampled a group of high-level public servants (Study 3A) and U.S. city mayors (Study 3B).

Methods For Study 3A, we recruited 107 high-level bureaucratic leaders in state and local government (73% male, mean age = 46 years, range: 25–63 years) enrolled in a multi-week on-site public policy executive education course at the Harvard Kennedy School.²⁶ Respondents were sent an email with a link to complete the study online (no payment was given for their participation). In our sample, 61% of respondents reported being elected or appointed to public office and 76% reported having the authority to directly affect public policy.

Our policymakers completed a survey in which they read about the strategic use of automatic enrollment defaults that was illustrated with either a liberal or conservative policy objective. Half of the policymakers read an example in which low-income earners were defaulted to automatically receive supplemental nutrition assistance benefits, and the other half read an example in which high-income earners were defaulted to automatically receive capital gains tax benefits. As in our previous studies, we reminded participants that they were evaluating the use of the nudge as a general-purpose policy tool, and asked participants to evaluate policy tools using the same six questions used in Studies 1 and 2 (Cronbach's $\alpha = .88$). Next, participants reported their political orientation for social and economic issues, as well as their attitudes toward the policy objective used to illustrate the nudge, following the same procedure as our previous studies.

For Study 3B, we recruited 48 U.S. city mayors with an average constituency base of 108,000 citizens (range: 8,000–620,000 citizens). Respondents were approached at a U.S. Conference of Mayors event and were asked to complete a 1-page survey in return for a copy of Thaler and

Sunstein's book *Nudge: Improving Decisions About Health, Wealth, and Happiness*. The design was similar to Study 3A except that the policy nudge was illustrated by automatically enrolling public high school students into supplemental educational programs for either safe sex practices or intelligent design.²⁷

Results Echoing the results from Study 2, policymakers tended to conflate their attitudes toward the policy nudge with their attitudes toward the policy objective used to illustrate it.²⁸ As displayed in the bottom two panels of Figures 1, liberal policymakers tended to exhibit greater support for automatic enrollment defaults when they were illustrated with liberal policy applications than conservative policy applications, and vice versa for conservative policymakers.

In Study 3A, we observe a significant interaction between political orientation and policy illustration (p = .014). When defaults were illustrated with automatic enrollment of low-income individuals into a food stamp program, conservative public servants were more likely than their liberal counterparts to oppose the use of defaults as general-purpose tools ($\beta = .38$, p < .001). In contrast, when defaults were illustrated with automatic enrollment of high-income citizens into a tax waiver program this effect did not appear ($\beta = .02$, p = .59). As in Study 2, we observed a positive relationship between attitudes towards the illustrated policy objective and evaluations of the policy nudge²⁹ ($\beta = .12$, p = .12).

Likewise, in Study 3B, we observe a similar interaction between political orientation and policy illustration (p = .078). When defaults were illustrated with automatic enrollment of high school students into intelligent design educational programs, liberal mayors were more likely than conservative mayors to oppose their use as general policy tools ($\beta = .44$, p = .042). In contrast, when defaults were illustrated with enrollment of students into safe sex educational programs this effect did not appear ($\beta = .05$, p = .64). Again, we observed a positive relationship between attitudes towards the illustrated policy objective and evaluations of the policy nudge ($\beta = .24$, p = .05).

General Discussion

We find that both laypeople and practicing policymakers evaluate policy nudges in ways that are colored by their political attitudes. People tend to view nudges as more unethical, coercive, and manipulative when enforced by a policymaker they oppose compared to one they support (Study 1), or when illustrated using policy objectives they oppose compared to objectives they support (Studies 2). Experienced policymakers also exhibit partisan nudge bias (Studies 3A and 3B).

Our findings are consistent with past psychological research documenting the difficulty people have in disentangling their attitudes towards a target object. For instance, people tend to be skeptical of policies or proposals from sources they distrust (analogous to what we observe in Study 1), and as a result devalue offers and concessions made by an adversary compared to when those same offers are made by an ally.³⁰ Likewise, past research in other domains has found that

people sometimes conflate their attitudes about inputs with the attitudes about outputs (analogous to what as we observed in Studies 2 and 3). For example, it is well known that people often confuse good or bad decision outcomes with good or bad decision reasoning,³¹ even though bad decisions sometimes result in positive outcomes (due to good luck) and good decision sometimes result in negative outcomes (due to bad luck). Similarly, researchers have recently found that ethical evaluations of new technologies (e.g., detecting discrimination through brain imaging techniques) are strongly affected by whether the discrimination in question accords with a respondent's political orientation (rooting out racism vs rooting out anti-Americanism).³²

It is worth emphasizing our finding that removing the specific policy context eliminated partisan nudge bias. In Studies 1 and 2 we administered control conditions that provided information about the policy nudge but omitted details about the policymaker or political objective tied to the policy nudge. When participants were placed behind this "veil of ignorance," both liberals and conservatives endorsed policy nudges to roughly the same extent, suggesting that attitudes towards nudges are not inherently partisan.³³ Furthermore, the range of policy nudges that were presented to participants in a decontextualized way (from the control condition in Study 2) tended to be rated positively (i.e. above the midpoint of the rating scale).³⁴ This finding is consistent with past surveys suggesting that the public is generally receptive towards policy nudges, especially when compared to more traditional forms of regulation (e.g., imposing taxes).^{35,36,37}

Naturally, our study leaves some important questions unanswered. First, our non-expert samples in the first two studies were not perfectly representative of the U.S. population so one might wonder the extent which our results apply more generally. However, we note that the subject pool used for these studies was more educated and politically knowledgeable than the U.S. populace.³⁸ Thus, we suspect that our results may, in fact, underestimate the magnitude of a partisan nudge bias. Second, our studies did not examine the degree to which ethical evaluations towards nudges translate into actionable behavior (e.g., voting), which would be a valuable direction for future research.³⁹ Third, we deliberately used descriptions of policies that were brief and incomplete; the intricacies and details of any public policy will be considerably more complex than what we presented to our participants. We chose brief descriptions to roughly mirror the amount of policy information from a standard news source. However, it is an open question whether more information exacerbates or attenuates partisan nudge bias.

Our findings suggest several policy prescriptions. First, advocates of behavioral policy interventions would be well-served to emphasize that such techniques are designed to improve the effectiveness of a policy objective that has been acknowledged to be a worthy priority. Shifting discussion to the capacity for nudges to enhance already-settled policy objectives sidesteps confusion between policy means and ends that is reflected in partisan nudge bias. Second, in light of our findings, it may be wise to establish general-purpose guidelines under which policy nudges are deemed acceptable or unacceptable, independent of any particular policy objectives (for example: Is the use of the nudge transparent? Does the nudge affect those who need guidance more strongly than those who need less guidance?). A recent successful example of this approach is the 2011 U.K.

House of Lords Science and Technology committee report on behavior change, ⁴⁰ which outlined a set of general criteria for evaluating the ethical acceptability of implementing policy nudges.

There may be valid reasons why policymakers choose to embrace or reject the application of behavioral insights to public policy, but a partisan nudge bias should not be one of them. The studies we report here suggest that policymakers and the public may prematurely dismiss effective policy tools, or embrace ineffective ones, for the wrong reasons. Such uncritical rejection and acceptance does a disservice to the critics and advocates of policy nudges who have substantive concerns about their merits. More importantly, it does a disservice to the public because governing institutions have a responsibility to carefully consider the range of policy tools at their disposal. Disentangling policy means from policy ends helps to focus political discourse where it belongs — on policy priorities — and also clarifies the terms of the debate about the legitimacy of using behavioral policy tools to advance those priorities.

Notes

¹Thaler, R. H. and Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press, New Haven, CT

²Madrian, B. C. and Shea, D. F. (2001). The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior. *Quarterly Journal of Economics*, 116(4):1149–1187

³Benartzi, S. and Thaler, R. H. (2013). Behavioral economics and the retirement savings crisis. *Science*, 339:1152–1153

⁴Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T., and Rabin, M. (2003). Regulation for conservatives: Behavioral economics and the case for "Asymmetric Paternalism". *University of Pennsylvania Law Review*, 151(3):1211–1254

⁵http://www.theguardian.com/commentisfree/2014/feb/05/nudge-say-no-more-behavioural-insights-team

⁶http://www.trn1.com/the-monica-crowley-show-government-think-tank-wants-to-nudge-voter-reaction

⁷http://mediamatters.org/embed/clips/2013/07/31/31291/fbc-dobbs-20130730-dobbsnudge

⁸The argument that reactions to brief sketches serve as an approximation for more reflective and deliberative attitudes was also made by Daniel Kahneman and Illana Ritov for the examination of contingent valuation methods: Kahneman, D. and Ritov, I. (1994). Determinants of stated willingness to pay for public goods: A study in the headline method. *Journal of Risk and Uncertainty*, 9(1):5–37

⁹We excluded 75 additional participants who had reported familiarity with the law from the analyses (we asked about familiarity with the PPA at the end of the study). Doing so allows for a clean test of our "control" condition in which information about the PPA was omitted. We had two additional screening criteria, which we enforced across studies. First, for duplicate IP addresses from online studies, we only included data from the first set of responses in order to avoid duplicate responding. This removed 2 surplus response sets from Study 1, 7 surplus response sets from Study 2, and 4 surplus response sets from Study 3A. Second, when participants reported their social and economic political orientation they had the option of instead indicating "completely unsure" or "hadn't given it much thought." Our primary analysis use political orientation as a predictor variable, so we removed participants who could not identify

their political orientation. This excluded an additional 24 participants from Study 1, 36 participants from Study 2, 3 participants in Study 3A, and 0 participants in Study 3B. The greater exclusion rate in Studies 1 and 2 likely reflects the population characteristics of our subject pool (Mechanical Turk), which tends to be younger on average and less politically involved than policymakers in Studies 3A and 3B. In the main text we report sample sizes and descriptive statistics after making these exclusions.

- ¹⁰One participant in Study 1 reported an age of 267 years. We interpreted this as a typo, and omitted this response when calculating age statistics of the sample.
- ¹¹For all studies conducted on MTurk, we used software that excluded participants who had completed any of the previous experiments. See Goldin, G., & Darlow, A. (2013). TurkGate (Version 0.4. 0)[Software].
- ¹²Paolacci, G., Chandler, J., and Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making*, 5(5):411–419
- ¹³Following common disclosure guidelines we have provided the full list of stimuli and measures for this study, as well as all subsequent studies, in the Supplemental Material. We also provide detailed descriptive statistics on the characteristics of our samples in the Supplemental Material.
- ¹⁴In Study 1 we estimated the following linear model:

$$NudgeEvaluations_{ij} = \alpha + \beta_1 Political_i + \beta_2 Policymaker_i + \beta_3 Political_i * Policymaker_i + \epsilon_{ij}$$
 (1)

where $NudgeEvaluations_{ij}$ represent evaluations by participant i in treatment condition j; $Policymaker_j$ is a vector of indicator variables for treatment condition j, with the control condition set as the base value; $Political_i$ is the self-reported political orientation for participant i, and takes a value of 1 to 7; and $Political_i * Policymaker_j$ is a vector of interaction terms that model the change in the slope of $Political_i$ as a function of treatment condition j. From the model we then calculate the average marginal effects (i.e., "simple slopes") of political orientation for each treatment condition, and perform pairwise comparisons. For this study and all subsequent studies we analyze results using heteroskedasticity-robust standard errors, and report results using regression weights that have been standardized within each sample (i.e., represent the standard-deviation change in nudge evaluations per standard deviation increase in political orientation).

- ¹⁵Nickerson, D. W. and Rogers, T. (2010). Do you have a voting plan? implementation intentions, voter turnout, and organic plan making. *Psychological Science*, 21(2):194–199
- ¹⁶Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C. (2011). Using implementation intentions prompts to enhance influenza vaccination rates. *Proceedings of the National Academy of Sciences*, 108(26):10415–10420
- ¹⁷Keller, P. A., Harlam, B., Loewenstein, G., and Volpp, K. G. (2011). Enhanced active choice: A new method to motivate behavior change. *Journal of Consumer Psychology*, 21(4):376–383
- ¹⁸Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, 47(2):263–291
- ¹⁹Rogers, T., Milkman, K. L., and Volpp, K. G. (2014). Commitment devices: using initiatives to change behavior. *JAMA*, 311(20):2065–2066
- ²⁰Dickerson, C. A., Thibodeau, R., Aronson, E., and Miller, D. (2003). Using cognitive dissonance to encourage water conservation. *Readings About The Social Animal*, page 283
- ²¹Goldstein, N. J. and Cialdini, R. B. (2007). Using social norms as a lever of social influence. In Pratkanis, A. R., editor, *The Science of Social Influence: Advances and Future Progress*, pages 167–191. Psychology Press, New York, NY

In particular, we estimated the following model:

$$NudgeEvaluations_{ijk} = \alpha + \beta_1 Political_i + \beta_2 Nudge_i + \beta_3 Illustration_k + \beta_4 Political_i * Illustration_k + \gamma_i + \zeta_i + \epsilon_{ijk}$$
 (2)

where $NudgeEvaluations_{ijk}$ represent evaluations by participant i for nudge j applied to policy goal k; $Political_i$ is the self-reported political orientation for participant i, and takes a value between 1 and 7; $Nudge_j$ is a vector of indicator variables for the particular policy nudge (the "automatic enrollment" condition was set as the base variable), and $Illustration_k$ is a vector of indicator variables for the particular policy goal attached to the nudge (the control condition was set as the base variable); and $Political_i * Illustration_k$ is a vector of interaction terms that models the change in the slope for $Political_i$ as a function being illustrated by policy application k. We also include participant and policy nudge random effects (γ_i and ζ_j , respectively), and ϵ_{ijk} represents residual error. After estimating the model we then calculated the average marginal effects (i.e., "simple slopes") of political orientation for each policy application k. It is these average marginal effects that we report in the main text above.

²⁵For each participant, we took the difference in their evaluations for nudges illustrated by conservative versus liberal policy applications (i.e., evaluations for tax breaks and intelligent design minus evaluations for food stamps and safe sex education). We then regressed this difference score onto participants political orientation. Positive coefficients imply that political conservatives would favor conservative applications and liberals would favor liberal applications, a negative coefficient would imply the opposite, and a coefficient of 0 implies no difference across evaluations as a function of political orientation.

²⁶Participants in Study 3A completed the survey before the executive education course began, and the course was not advertised as focusing on behavioral policy interventions. During the course participants were exposed to approximately three hours of curriculum on behavioral policy interventions.

²⁷Also different from Study 3A was that respondents responded to a shorter set of evaluation items (how much they supported the nudge, opposed the nudge, found it manipulative, and found it coercive; Cronbach's $\alpha = .89$) and instead reported their general political orientation using a single-item scale.

²⁸Studies 3A and 3B implemented a similar analysis strategy to Study 1 (see Endnote 14), with the only difference being that these models included an indicator variable for illustrated policy application (0 = liberal objective, 1 = conservative objective) instead of for policymaker. Importantly, the sample sizes of Studies 3A and 3B were relatively small, so we took some additional precautions. Conventional (parametric) approaches rely on the large-sample properties of test statistics, and therefore may be inappropriate for our analysis. To address this problem we calculated one-tailed *p*-values using exact permutation tests, which are independent of sample size. We randomly reshuffled nudge evaluations 10,000 times; in each case we calculated the *t*-statistic for the interaction term of political orientation and nudge illustration, as well *t*-statistics for the simple slopes (i.e., the average marginal effect of political orientation within each nudge illustration). Exact (one-tailed) p-values were then calculated by comparing permutations to observed values. We justify the use of

²²Allcott, H. and Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, 104(10):1–37

²³Kahan, D. M., Jenkins-Smith, H., and Braman, D. (2011). Cultural cognition of scientific consensus. *Journal of Risk Research*, 14(2):147–174

²⁴Participants responded to multiple scenarios, so we analyzed responses using a mixed-effect linear model with crossed random effects for participants and policy nudges. For an overview, see: Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4):390–412; Judd, C. M., Westfall, J., and Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *Journal of personality and social psychology*, 103(1):54.

- one-tailed tests based on the consistent pattern of findings we observed in the previous studies. As a robustness check we also conducted (large-sample) bootstrapping procedures and found similar results.
- ²⁹The weak relationship between policy objective attitudes and nudge evaluations in Study 3A masks an unexpected difference between conditions, which mirrors the interaction between policy illustration and political orientation. We find a sizable effect when examining attitudes toward food assistance programs on nudge evaluations ($\beta = .46$, p < .001), whereas no reliable effect when examining attitudes toward high-income tax breaks ($\beta = .00$, p = .99). The size of this relationship across the two policy illustrations was also significant (p = .006 for the interaction term).
- ³⁰Ross, L. (1995). Reactive devaluation in negotiation and conflict resolution. In et al., K. A., editor, *Barriers to the negotiated resolution of conflict*, pages 30–48. Norton, New York
- ³¹Baron, J. and Hersjey, J. C. (1988). Outcome bias in decision evaluation. *Journal of personality and social psychology*, 54(4):569–579
- ³²Tetlock, P. E., Mitchell, G., and Anastasopoulos, L. J. (2013). Detecting and punishing unconscious bias. *The Journal of Legal Studies*, 42(1):83–110
- ³³For a similar finding involving international negotiations on the economic burdens of solving climate change see: Kriss, P. H., Loewenstein, G., Wang, X., Weber, R. A., et al. (2011). Behind the veil of ignorance: Self-serving bias in climate change negotiations. *Judgment and Decision Making*, 6(7):602–615
- 34 Restricting the analysis to the control condition, three of the five nudges were reliably above the midpoint at p < .10 (implementation intentions, commitment and consistency, highlighting losses), one nudge was reliably below the midpoint (automatic enrollment defaults), and one was not reliably different from the midpoint (descriptive social norms). Descriptive and inferential statistics are provided in the Supplemental Material.
- ³⁵Attari, S. Z., Schoen, M., Davidson, C. I., DeKay, M. L., de Bruin, W. B., Dawes, R., and Small, M. J. (2009). Preferences for change: Do individuals prefer voluntary actions, soft regulations, or hard regulations to decrease fossil fuel consumption? *Ecological Economics*, 68(6):1701–1710
- ³⁶Sunstein, Cass R. (May 10, 2015). Do People Like Nudges? Available at SSRN: http://ssrn.com/abstract=2604084
- ³⁷William Hagman et al., Public Views on Policies Involving Nudges, Review of Philosophy and Psychology (forthcoming 2015), available at http://www.iei.liu.se/nek/forskning/jedi-lab/1.630217/Nudge20150417.pdf
- ³⁸Berinsky, A. J., Huber, G. A., and Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon. com's mechanical turk. *Political Analysis*, 20(3):351–368
- ³⁹Another open question for future research is whether partisan nudge bias is drive more by "ingroup support" (i.e., insufficiently critical towards nudges when applied towards objectives one supports or by policymakers one supports) or by "outgroup opposition" (i.e., excessively critical towards nudges when applied towards objectives one opposes or by policymakers one opposes).
- ⁴⁰http://www.publications.parliament.uk/pa/ld201012/ldselect/ldsctech/179/179.pdf