

On the misplaced politics of behavioural policy interventions

David Tannenbaum^{1*} , Craig R. Fox² and Todd Rogers³

Government agencies around the world have begun to embrace the use of behavioural policy interventions (such as the strategic use of default options), which has inspired vigorous public discussion about the ethics of their use. Since any feasible policy requires some measure of public support, understanding when people find behavioural policy interventions acceptable is critical. We present experimental evidence for a 'partisan nudge bias' in both US adults and practising policymakers. Across a range of policy settings, people find the general use of behavioural interventions more ethical when illustrated by examples that accord with their politics, but view those same interventions as more unethical when illustrated by examples at odds with their politics. Importantly, these differences disappear when behavioural interventions are stripped of partisan cues, suggesting that acceptance of such policy tools is not an inherently partisan issue. Our results suggest that opposition to (or support for) behavioural policy interventions should not always be taken at face value, as people appear to conflate their attitudes about general purpose policy methods with their attitudes about specific policy objectives or policy sponsors.

A primary objective of public policy is to enhance welfare by shaping behaviour. To do this, policymakers have traditionally relied on a limited set of tools that include taxes, subsidies, mandates, bans and information campaigns. Such techniques are predicated on the assumption from neoclassical economics that individuals and organizations act according to their rational self-interest. More recently, government agencies around the world have begun to add behavioural policy interventions to their toolkits¹⁻³. For instance, in 2015, then-president Barack Obama directed federal agencies by executive order to identify and test applications of behavioural insights to improve the effectiveness of their programmes^{4,5}. Behavioural policy interventions leverage insights from the social and behavioural sciences, which paint a more complete portrait of human information processing, motivation and decision-making. A prominent form of behavioural policy interventions, which we focus on in this paper, are interventions that 'nudge' desired behaviours without meaningfully altering material incentives or limiting freedom of choice. Examples of nudges include providing citizens with information about how their home energy use compares with that of their neighbours and encouraging employees to participate in 401(k) retirement savings plans through the use of automatic enrollment defaults^{6,7}.

Behavioural policy interventions represent a new set of tools for making public policy more effective and cost-efficient^{8,9}. At the same time, there has been vigorous public discussion about the acceptability of using such techniques, with opponents fearing that

government nudges are manipulative and coercive¹⁰. When the relatively conservative administration of former UK Prime Minister David Cameron began applying behavioural insights to public policy, concerns of coercion came mostly from the political left. An article in the *The Guardian*, for instance, ran with the headline 'Nudge nudge, say no more. Brits' minds will be controlled without us knowing it'¹¹. In the United States, however, the creation of a similar behavioural insights agency under the relatively liberal administration of former president Barack Obama instead met skepticism primarily from the political right. One commentator on *Fox News*, for instance, characterized the effort as 'propaganda and mind control'¹². These anecdotes suggest that people express concern about the ethical use of behavioural policy interventions, but that such reservations may not fall neatly along traditional ideological lines.

Understanding how people feel about the use of nudges is critical because their implementation by policymakers requires some measure of public support¹³. Recent surveys in a number of countries have found general support for nudges and other behavioural policy interventions, especially when compared with conventional policy tools such as taxes or mandates¹⁴⁻¹⁹. This research suggests that people have discernible preferences for different policy approaches, and tend to prefer techniques that do not impinge on individual autonomy or limit freedom of choice²⁰. These surveys also find that political ideology tends to be a weak predictor of general support for behavioural policy interventions^{17,19}. At the same time, well-established research literature in both psychology²¹⁻²³ and political science²⁴⁻²⁷ has found that support for policies is often coloured by partisan affiliations. For instance, people are more supportive of a proposed welfare reform policy when informed that the policy is endorsed rather than opposed by their preferred political party²⁸. However, most of this research has focused on support for specific policy goals for which individuals tend to have well-established preferences (for example, welfare policy). It is an open question whether partisan cues distort attitudes towards general purpose nudges, which are explicitly described as issue-agnostic and contain features that most people see as inherently desirable (for example, preserving freedom of choice).

Here, we examine whether partisans from both ends of the political spectrum conflate their feelings about the use of general-purpose policy nudges with their feelings about salient examples of specific policy objectives to which those interventions might be applied or their feelings about the policymakers who endorse such tools. We refer to this phenomenon as 'partisan nudge bias'. Partisan nudge bias can be thought of as an instance of attribute substitution^{29,30}, where individuals evaluate the acceptability of a policy nudge by instead assessing how they feel about the associated policy objective or policy sponsor.

¹Eccles School of Business, University of Utah, Salt Lake City, Utah 84112, USA. ²Anderson School of Management, University of California, Los Angeles, Los Angeles, California 90095, USA. ³Kennedy School of Government, Harvard University, Cambridge, Massachusetts 02138, USA.

*e-mail: david.tannenbaum@utah.edu

We conducted four experiments testing whether US citizens and practising policymakers exhibit partisan nudge bias. Study 1 was designed to examine partisan nudge bias across a range of policy nudges and policy settings. We presented a sample of 462 US adults with short descriptions of five empirically validated, field-tested policy nudges. Each policy nudge was shown on a separate page and explained in nonpartisan language. Importantly, we randomly paired each nudge with an illustration of how the technique could be applied to either a politically conservative or liberal policy goal. For instance, for one policy nudge, participants learned that policymakers can often increase participation in a programme by providing information about how other people behave in the same situation (that is, descriptive social norms). From there, some participants read an illustration of using descriptive social norms to increase the uptake of tax breaks for high-income citizens (a relatively conservative policy goal), whereas other participants read an illustration of using descriptive norms to increase the uptake of supplemental nutrition assistance programmes for low-income citizens (a relatively liberal policy goal). The pairing of illustrations to nudges was randomized for each participant, with the constraint that participants viewed each nudge and illustration once and only once.

After reading each policy nudge and accompanying illustration, participants rated the acceptability of the nudge. We first reminded participants that the policy technique could be deployed “across a wide range of policies beyond the illustration above” and then asked participants, “Putting this particular example aside, how do you feel about [policy nudge] as a general approach to public policy?” (emphasis in original). Participants responded on five-point scales indicating how much they supported general use of the technique, how much they opposed its general use, as well as how ethical, unethical, coercive and manipulative they found its use to be (for example, 1 = strongly oppose, 5 = strongly support). We combined these items to form an index of nudge attitudes, with higher numbers indicating greater endorsement. At the end of the survey, participants reported their political orientation for economic and social issues.

The design of study 1 provides a direct test of partisan nudge bias, since descriptions of policy nudges were randomly paired with partisan policy illustrations. Across the range of nudges sampled, and consistent with partisan nudge bias, liberal respondents reported greater acceptance of the general use of policy nudges when illustrated by liberal (as opposed to conservative) policy objectives; meanwhile, conservative respondents reported greater acceptance when nudges were illustrated by conservative (as opposed to liberal) policy objectives ($P < 0.001$ for the interaction between respondents’ political orientation and policy illustrations; see Methods for a detailed description of the analysis strategy for all the studies). To provide a sense of the magnitude of partisan nudge bias, we estimate the degree of bias (that is, differences in acceptance when illustrated by liberal versus conservative policy goals) relative to the overall distribution in nudge attitudes across all treatment conditions. As depicted in Fig. 1a, the most partisan respondents (top and bottom bars) tended to be more biased while politically moderate respondents (bars closer to the middle) tended to be less biased. Participants reporting a ‘very liberal’ orientation would be expected to show a bias in favour of liberal policy applications equivalent to a 0.62 s.d. in nudge attitudes, whereas participants reporting a ‘very conservative’ orientation would be expected to show the opposite bias corresponding to a 0.26 s.d. in nudge attitudes. Inspecting each policy nudge separately, the interaction term is in the predicted direction for all five nudges (Fig. 2a–e). Furthermore, in study 1 we were also able to conduct a within-participant analysis, which holds all unobservable participant characteristics fixed. Again, we found evidence of partisan nudge bias (Pearson’s correlation coefficient, $r = 0.24$, $P < 0.001$).

In study 2, we examined whether partisan nudge bias emerges when both the nudge and policy illustration are held constant, while

information about the identity of the policy sponsor (either a prominent liberal or a prominent conservative) is varied. Such a finding would suggest that individuals are not merely responding to associations with particular policy illustrations, but rather exhibiting more general partisan reactions. A sample of 355 US adults read about the use of automatic enrollment defaults as a general policy tool, and that the US Congress had recently passed a law called the Pension Protection Act (PPA), which encouraged companies to automatically enroll their employees in retirement savings plans while allowing them to opt out. We focused on the use of automatic enrollment because it is the best-known and most widely used behavioural policy intervention. Importantly, participants were randomly assigned to learn that the PPA was enforced by either the Bush or Obama administration (in reality, both administrations enforced the law, which passed with strong bipartisan support in 2006). Participants then rated the acceptability of automatic enrollment as a general policy tool using the same measures as in study 1.

Participants again exhibited partisan nudge bias, with support for the use of automatic enrollment depending on which administration participants learned had enforced the policy. Figure 1b shows that liberal participants were more likely to find automatic enrollment unacceptable when informed that the Bush administration (as opposed to the Obama administration) had enforced the PPA. Conversely, conservative participants exhibited the reverse pattern of results ($P = 0.05$ for the interaction between respondents’ political orientations and information about the policy sponsor). Based on our results, ‘very liberal’ participants would be expected to show a liberal partisan nudge bias corresponding to a 0.51 s.d. in nudge attitudes, whereas ‘very conservative’ participants would be expected to show the opposite bias corresponding to a 0.60 s.d. in nudge attitudes.

Studies 1 and 2 document partisan nudge bias among US adults. However, this leaves open the question of whether experienced policymakers are also susceptible to partisan nudge bias. For study 3, we recruited senior-level state and local government leaders ($n = 107$) and for study 4, we recruited US city mayors with an average constituency base of 108,000 citizens ($n = 48$). Policymakers completed a survey in which they read about the use of automatic enrollment defaults illustrated by either a liberal or conservative policy objective (similar to study 1). As in our previous studies, we reminded these policymakers that their task was to evaluate the general use of the policy nudge rather than a specific implementation. They then rated the acceptability of automatic enrollment using questions similar to those used in our earlier studies.

Studies 3 and 4 found that both senior government leaders and US mayors exhibit partisan nudge bias, just as our general sample of adults did in studies 1 and 2. As displayed in Fig. 1c,d, liberal policymakers exhibited greater support for automatic enrollment when paired with liberal (as opposed to conservative) policy applications, while conservative policymakers showed greater support for automatic enrollment when paired with conservative (as opposed to liberal) policy applications. The significance of the interaction between political orientation and policy illustrations was $P = 0.017$ for study 3 and $P = 0.079$ for study 4. Based on our results, ‘very liberal’ policymakers in study 3 would be expected to show a liberal partisan nudge bias corresponding to a 0.73 s.d. in nudge attitudes, whereas ‘very conservative’ policymakers would be expected to show a conservative partisan nudge bias corresponding to a 0.85 s.d. in nudge attitudes. For study 4, ‘very liberal’ policymakers would be expected to show a liberal partisan nudge bias corresponding to a 0.86 s.d. in nudge attitudes, whereas ‘very conservative’ policymakers would be expected to show a conservative partisan nudge bias corresponding to a 1.00 s.d. in nudge attitudes.

It is possible that partisan nudge bias operates as an instance of attribute substitution^{29,30}, where individuals evaluate the acceptability of a policy nudge by instead assessing how they feel about the

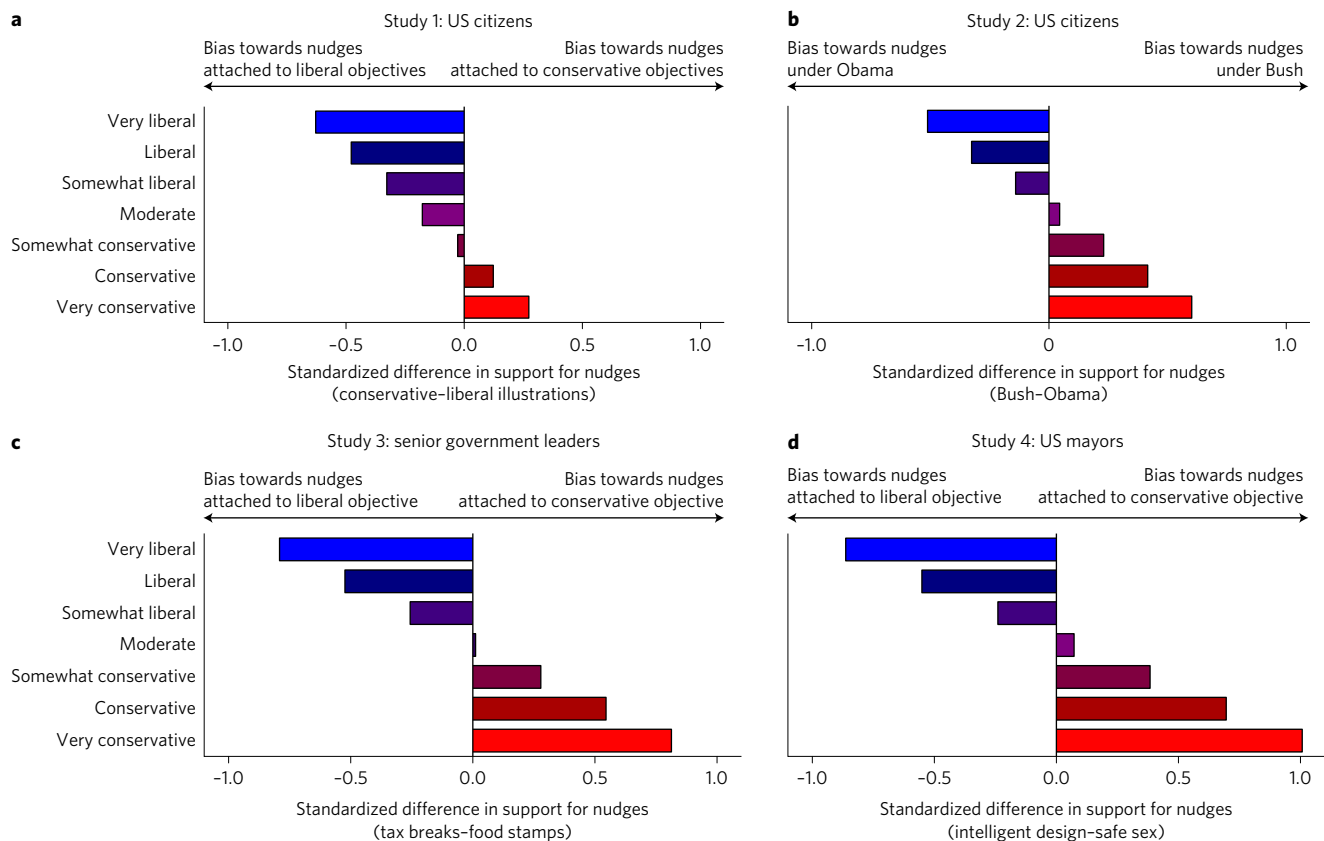


Figure 1 | Expected differences in nudge attitudes based on the regression specification for each study. The y axes indicate participant political orientation, whereas scores on the x axes represent standardized difference scores in nudge attitudes (relative to the overall distribution in nudge attitudes across treatment conditions). **a**, For study 1 ($n=462$), the difference scores represent nudge attitudes when illustrated by liberal versus conservative policy objectives (with evaluations aggregated across all policy nudges). **b**, For study 2 ($n=355$), the difference scores represent nudge attitudes under Bush versus Obama. **c**, For study 3 ($n=107$), the difference scores represent nudge attitudes when illustrated by tax breaks versus supplemental nutrition assistance. **d**, For study 4 ($n=48$), the difference scores represent nudge attitudes when illustrated by intelligent design versus safe sex.

associated policy objective or policy sponsor. To examine this possibility, at the end of each study, we asked participants to report their attitudes towards the policy goals that served as illustrations (in studies 1, 3 and 4) or their attitudes towards the policy sponsor (in study 2). Consistent with the idea of partisan nudge bias as attribute substitution, attitudes towards policy nudges tended to be reliably associated with attitudes towards the associated policy objective or sponsor (combined $r=0.29$, $P<0.001$; see Supplementary Table 2).

Collectively, studies 1 to 4 support the notion that both laypeople and policymakers evaluate policy nudges in ways that are coloured by their partisan affiliation. This suggests that stripping away partisan cues from the descriptions of nudges should attenuate, if not expunge, partisan nudge bias. To test this hypothesis, in studies 1 and 2 we included additional ‘control’ conditions where policy nudges were described omitting specific policy illustrations or sponsor information. In both studies, partisan nudge bias disappeared entirely: participants’ political leanings were not reliably related to nudge attitudes (combined $r=0.00$ for study 1; $r=-0.06$ for study 2; $P\geq 0.50$ for both studies). Furthermore, participants generally tended to rate policy nudges positively (that is, above the midpoint of the scale) across the range of descriptions in the control conditions of study 1 (see Supplementary Table 3).

Naturally, people may embrace or reject policy nudges for principled reasons, such as preferences for or against government intervention (that is, libertarian sensibilities). In fact, at the end of study 1, we also measured general attitudes towards government intervention using a scale of libertarianism²³. We found that more libertarian

participants tended to rate all nudges as less acceptable when no policy illustrations were provided (that is, in the control conditions; combined $r=-0.22$, $P<0.001$). However, the size of partisan nudge bias was roughly threefold greater (in absolute magnitude) than that of libertarian objections to policy nudges. In the non-control treatment conditions—where participants were provided with concrete illustrations of behavioural interventions—nudge attitudes were more strongly predicted by their attitudes to the associated policy objective (that is, partisan nudge bias; combined $r=0.30$, $P<0.001$) than by libertarian sensibilities (combined $r=-0.10$, $P=0.003$).

In summary, we observed that both laypeople and practising policymakers evaluate policy nudges in ways that are coloured by their political preferences. People tend to view nudges as more unethical, coercive and manipulative when illustrated by policy objectives they oppose compared with objectives they support (study 1), or when told that such behavioural interventions have been enforced by a policymaker they oppose compared with one they support (study 2). Experienced policymakers also exhibit partisan nudge bias (studies 3 and 4).

Importantly, when we described nudges without information about policy objectives or advocates that might provide partisan cues, liberals and conservatives evaluated these interventions similarly. This finding suggests that there may not be strong partisan objections to the use of policy nudges. Thus, we recommend that, whenever feasible, policymakers should minimize partisan information associated with a policy nudge. This recommendation is consistent with previous research, which suggests that the most

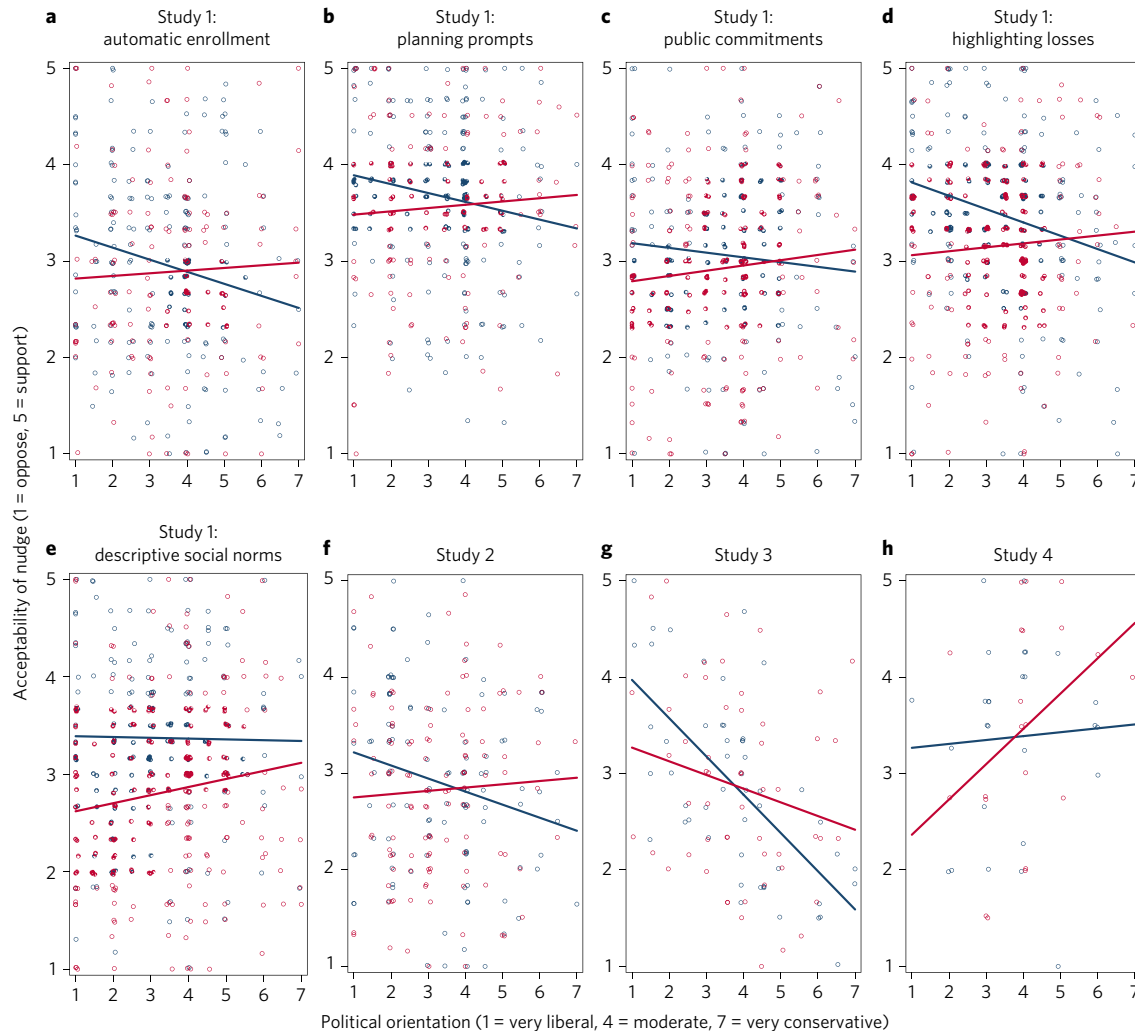


Figure 2 | Graphical overview of studies 1–4. Greater numbers on the y axes indicate more positive nudge attitudes, whereas greater numbers on the x axes indicate more conservative political orientation. Data points have a small amount of ‘jitter’ added to indicate density. **a–e**, For study 1 ($n = 462$), the blue and red data points represent nudge attitudes illustrated by liberal and conservative policy applications, respectively, in relation to automatic enrollment (**a**), planning prompts (**b**), public commitments (**c**), highlighting losses (**d**) and descriptive social norms (**e**). **f**, For study 2 ($n = 355$), the blue and red data points represent evaluations when the policy sponsor was the Obama or Bush administration, respectively. **g**, For study 3 ($n = 107$), the blue and red data points represent nudge attitudes illustrated by applications to supplemental nutrition assistance or tax breaks, respectively. **h**, For study 4 ($n = 48$), the blue and red data points represent nudge attitudes illustrated by applications to safe sex education or intelligent design education programmes, respectively. The lines represent the best fit for each condition.

effective way to reduce bias is to prevent the inputs leading to bias from being triggered in the first place^{31,32}. That said, we note that removing partisan cues from a policy proposal is often impractical or untenable (for example, information on the sponsor of a particular bill cannot be removed). Thus, advocates of behavioural policy interventions may be better served by emphasizing that such techniques are designed to improve the efficiency of programmes that address acknowledged policy priorities (for example, increase retirement savings or conserve energy). Shifting discussion to the capacity for policy nudges to enhance already settled policy objectives avoids confusion between policy means and ends that are reflected in partisan nudge bias.

In a similar vein, policymakers could employ multiple illustrations from across the political spectrum when discussing the appropriateness of a particular nudge. Allowing individuals to jointly compare different possible applications of nudges can serve as an effective debiasing technique because such side-by-side comparisons make it easy for individuals to discount biasing information³³.

Highlighting that behavioural interventions can be applied to both liberal and conservative policy goals should focus individuals on the appropriateness of such general purpose techniques, rather than on a particular application of that technique.

Finally, in the light of our findings, it may be wise to establish general-purpose guidelines under which policy interventions are deemed acceptable or unacceptable, independent of any particular policy objective (for example, ‘Is the use of the nudge transparent?’ or ‘Does the intervention affect those who need more guidance more strongly than those who need less guidance?’). A successful example of this approach is the 2011 UK House of Lords Science and Technology Committee report on behaviour change³⁴, which outlined a set of general criteria for evaluating the ethical acceptability of implementing behavioural policy techniques.

It is also an open question whether nudges, compared with other policy tools, are particularly susceptible to partisan bias. One possibility is that people do not have clear attitudes towards any policy interventions (whether they are behavioural or conventional in

nature), and so rely on partisan cues to formulate evaluations of all policy techniques. However, previous research suggests that this is not the case, as people have clear preferences for many kinds of policy interventions independent of how they are applied. For instance, surveys find that people tend to hold considerably more negative views of mandates and bans than behavioural policy interventions such as automatic enrollment defaults, even when the mandates and bans are applied to more acceptable policy ends than behavioural policy interventions²⁰. While further research is needed, behavioural interventions may be uniquely susceptible to partisan bias because of how they influence citizens' behaviour. Behavioural interventions (for example, changing a default option) often operate covertly (that is, people are not aware of a nudge's influence on their decisions), whereas conventional policy interventions (for example, offering a tax credit or penalty) are typically more amenable to conscious deliberation and, therefore, volition. As such, behavioural interventions may be more open to charges of manipulation and coercion by those who are skeptical of the policymakers' aims or of the policy agenda to which the nudge is applied.

Our findings underscore the difficulty policymakers will likely face when employing behavioural policy tools in a politically polarized environment. Citizens and policymakers may prematurely reject or accept powerful new strategies for achieving policy objectives merely because they dislike or like the policies with which these strategies happen to be associated. Considering the acceptability of policy nudges independent of potentially partisan ends, or at least illustrated by noncontroversial policy goals, should lead to more thoughtful debate about when such tools are ethically appropriate.

Methods

For each study, we predetermined our sample size and report all data exclusions, manipulations and measures³⁵. All study materials are available online in the Supplementary Information. All participants gave informed consent and the research was approved by the University of California, Los Angeles, and the University of Chicago Institutional Review Boards.

Study 1. Participants. We did not have a baseline measure of effect size, so we made an informal estimate of requisite power needs and set in advance a sample size target of 500 participants. This would give us a lower bound of 80% power to detect an effect size of $d = 0.13$. We recruited participants from Amazon.com's Mechanical Turk labour market to complete our online survey in return for a flat cash payment. We recruited 496 participants and, after data exclusions (discussed below), our final sample consisted of 462 participants (67% male, mean age: 32 years, range: 18–74 years).

Design. Each participant read short descriptions of five empirically established, field-tested policy nudges. These included the use of automatic enrollment defaults^{7,36,37} ('automatic enrollment'), prompting citizens to formulate concrete action plans^{38–40} ('planning prompts'), highlighting the potential losses associated with undesirable behaviour rather than highlighting the potential gains associated with desirable behaviour^{41,42} ('highlighting losses'), prompting citizens to publicly commit to behaviours in advance to encourage follow-through^{43,44} ('public commitments') and providing citizens with information about how other people behave in the same situation^{6,45,46} ('descriptive social norms').

Each nudge was randomly paired with an example illustration of its application to a particular policy objective. Illustrations of policy applications included increasing participation in: (1) supplemental nutrition assistance programmes by low-income individuals ('food stamps'), (2) tax break benefits on earned capital gains by high-income individuals ('tax breaks'), (3) public high school education programmes on safe sex and effective contraceptive use ('safe sex') and (4) public high school education programmes on intelligent design ('intelligent design'). These four policy applications were designed to appeal to economic liberals, economic conservatives, social liberals and social conservatives, respectively. Along with the four policy illustrations, we included a generic, context-free policy illustration ('control condition'). Each policy description was presented on a separate page, and we randomized the order in which nudges were presented as well as the example policy illustrations. Participants viewed each policy nudge and policy illustration once and only once; thus, individual participants responded to five policies from a pool of 25 possible pairings.

After reading each policy description (and accompanying illustration), participants evaluated the use of the policy nudge "as a general approach to public policy". We explicitly instructed participants to report their feelings for general use of the nudge, setting the particular application aside, and reminded participants

that the use of the behavioural policy tool could be applied "across a wide range of policies beyond the illustration above". For each policy description, participants responded to six questions asking how much they supported using the nudge as a general approach to public policy, opposed using the nudge as a general approach to public policy, and how ethical, unethical, coercive and manipulative they found their use to be (for example, 1 = strongly oppose, 5 = strongly support). All items were combined and averaged to form an index of policy nudge evaluations (Cronbach's α ranged from 0.87 to 0.89 for behavioural policy tools when pooling over illustrations), with higher numbers indicating greater endorsement of their use.

At the end of the study, we asked participants a series of demographic and attitude questions that included their political orientation. Participants rated their political orientation separately for economic and social issues (1 = very liberal, 7 = very conservative); for both items, participants also had the option of indicating that they were "completely unsure" of their political orientation or "haven't given it much thought". For the analyses reported in this paper, we averaged these two responses for each individual to create a single index of political orientation (Supplementary Table 4 reports all analyses when social and economic political orientation are examined separately). Participants also rated their attitudes towards each of the policy goals that served as illustrations (for example, "How do you feel about promoting food stamp programmes for low-income individuals?") on seven-point scales (1 = strongly oppose, 7 = strongly support). Lastly, participants completed a six-question scale that measured individual differences in libertarianism²³. For this scale, participants indicated their degree of agreement (1 = strongly disagree, 7 = strongly agree) with statements such as "It's not the government's business to try to protect people from themselves", and "Sometimes government needs to make laws that keep people from hurting themselves". All responses were combined, with higher scores indicating greater endorsement of libertarianism (Cronbach's $\alpha = 0.83$).

Data exclusions. Participants sometimes completed the survey more than once, so we dropped seven responses that indicated a duplicate IP address; including duplicate IP addresses in the analysis does not change the coefficient signs or significance levels of our results. We also excluded 34 participants who did not report an identifiable social or economic political orientation (that is, they reported being "completely unsure" or that they "haven't given it much thought"). We excluded these participants because having an identifiable political orientation was necessary to conduct our primary analyses (that is, political orientation was a key independent variable).

Analysis strategy. We predicted that more liberal participants would report greater support for behavioural policy tools when illustrated by liberal objectives rather than conservative objectives, while more conservative participants would show more support for tools illustrated by conservative objectives. To test this, we estimated the following ordinary least squares model:

$$\text{NudgeEvaluations}_{ijk} = \alpha + \beta_1 \text{Partisan}_i + \beta_2 \text{PolicyTool}_j + \beta_3 \text{Application}_k + \beta_4 \text{Partisan}_i \times \text{Application}_k + \epsilon_{ijk}$$

where $\text{NudgeEvaluations}_{ijk}$ represent evaluations by participant i for policy nudge j illustrated by policy application k , α represents the model intercept, and β the regression weights for each predictor variable. Partisan_i is the self-reported political orientation for participant i and takes a value between 1 and 7, with higher numbers reflecting greater conservatism; positive coefficients indicate that more conservative participants showed greater support for the use of the policy nudge. PolicyTool_j is a vector of indicator variables for each policy nudge. Application_k is an indicator taking a value of 1 if the policy tool was illustrated by a conservative policy goal ('tax breaks' or 'intelligent design') and a value of 0 if the policy tool was illustrated by a liberal policy goal ('food stamps' or 'safe sex'); positive coefficients indicate greater support for policy nudges illustrated with conservative rather than liberal policy applications. The interaction term indicates the change in Application_k as a function of participants' political orientation; a positive interaction term therefore provides support for partisan nudge bias. We implemented robust standard errors clustered by participants (ϵ_{ijk}). To provide an estimate of the size for the interaction effect, we conducted a 'spotlight' test, which provides an estimate of the treatment effect (that is, Application_k) at different values of political orientation⁴⁷, and then standardized these estimates over the pooled s.d. in nudge attitudes across the entire sample.

For our within-participant analysis, we calculated difference scores for each participant in relation to nudges illustrated by conservative versus liberal policy applications (positive scores indicate a bias in favour of nudges illustrated by conservative applications and negative scores indicate a bias in favour of liberal applications). We correlated participants' difference scores with their political orientation ratings; a positive correlation would provide support for partisan nudge bias.

For both of the analyses above, we excluded 'control' illustrations from the model and analysed those data separately. We did this for purposes of simplicity, as it makes the analysis less cumbersome; including the control conditions in the model does not change the coefficient signs or significance levels of any of our reported findings.

When aggregating correlations both within studies and across studies (for example, the relationship between attitudes about policy goals and nudges across our five policy illustrations), we used the sample size weighted mean correlation⁴⁸. Using the alternative method of first converting correlations to Fischer's *Z* before aggregation⁴⁹ returns virtually identical results.

Study 2. Participants. We made an informal estimate of requisite power needs and set in advance a sample size target of 450 participants (150 participants per condition). At 150 participants per condition, this would give us 80% power to detect an effect size of $d = 0.32$. We recruited participants from Amazon.com's Mechanical Turk labour market to complete our online survey in return for a flat cash payment. We recruited 455 participants and, after data exclusions, our final sample consisted of 355 participants (54% male, mean age: 30 years, range: 18–67 years).

Design. 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 US 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 Bush administration ($n = 119$), the Obama administration ($n = 116$), or a control condition where the PPA had simply been enforced by 'lawmakers' ($n = 120$). After reading about the policy, participants then responded to the same questions used in study 1. All items were combined and averaged to form an index of nudge evaluations (Cronbach's $\alpha = 0.89$), with higher numbers indicating greater endorsement of the general use of policy nudges. At the end of the study, participants reported their political orientation using the same items from study 1. Participants also rated the degree to which they found presidents Bush and Obama trustworthy (1 = not at all trustworthy, 7 = extremely trustworthy).

Data exclusions. Participants sometimes completed the survey more than once, so we dropped two responses that indicated a duplicate IP address; including duplicate IP addresses in the analysis does not change the coefficient signs or significance levels of our results. As with study 1, we also excluded 24 participants who did not report an identifiable political orientation (that is, they reported being "completely unsure" or that they "haven't given it much thought"). Lastly, we excluded 73 participants from the analysis who reported familiarity with the PPA. Unsurprisingly, these participants were not reliably affected by information about whether Obama or Bush had enforced the policy ($P = 0.66$ for the interaction term). Including these participants in the analysis therefore slightly dilutes our effect and increases the *P* value for the interaction term from 0.05 to 0.12.

Analysis strategy. We expected that partisans would conflate their attitudes towards the behavioural policy tool with their attitudes towards the mentioned policy sponsor. To test this, we estimated the following ordinary least squares model:

$$\begin{aligned} \text{NudgeEvaluations}_{ij} & \\ &= \alpha + \beta_1 \text{Partisan}_i + \beta_2 \text{PolicyMaker}_j \\ &+ \beta_3 \text{Partisan}_i \times \text{PolicyMaker}_j + \epsilon_{ij} \end{aligned}$$

where $\text{NudgeEvaluations}_{ij}$ represent attitudes towards automatic enrollment by participant i in treatment condition j . Partisan_i is the self-reported political orientation for participant i and takes on a value between 1 and 7, with higher numbers reflecting greater conservatism; a positive coefficient indicates that more conservative participants reported greater support for the use of the behavioural policy tool. PolicyMaker_j is an indicator taking a value of 1 if participants were informed that the policy was enforced by the Bush administration and a value of 0 if they were informed that the policy was enforced by the Obama administration; a positive coefficient indicates greater support for behavioural policy tools when participants learned Bush was the policy sponsor. The interaction term indicates the change in PolicyMaker_j as a function of participants' political orientation; a positive coefficient would provide support for partisan nudge bias. We implemented robust standard errors (ϵ_{ij}) to account for arbitrary forms of heteroskedasticity. To provide an estimate of the size for the interaction effect, we performed the same standardized 'spotlight' test as that used in study 1.

Study 3. Participants. We recruited a sample of high-level government leaders from a public policy executive education course at the Harvard Kennedy School. There was no sample size target; however, we aimed to include as many eligible participants as we could. We invited all executives who were enrolled in the course to participate in the study by sending them an email with a link to complete the study online (no payment was given for their participation). Participants completed the survey before the executive education course began, and the course was not advertised as focusing on behavioural policy interventions; during the course, participants were exposed to approximately three hours of curriculum on behavioural policy interventions. After data exclusions, our final sample consisted

of 107 participants (73% male, mean age: 46 years, range: 25–63 years). Most respondents reported that they were elected or appointed to public office (61%) and that they had the authority to directly affect public policy (76%).

Design. Participants read about the strategic use of automatic enrollment defaults illustrated with either a liberal or conservative policy objective. Policymakers were randomly assigned to read either an example in which low-income earners were defaulted to automatically receive supplemental nutrition assistance benefits ($n = 53$) or an example in which high-income earners were defaulted to automatically receive capital gains tax benefits ($n = 58$). Similar to our previous studies, we reminded participants that they were evaluating the general acceptability of using the behavioural policy tool, and then they responded to the same six questions used in studies 1 and 2 (Cronbach's $\alpha = 0.88$). Next, participants reported their political orientation for social and economic issues, as well as their attitudes towards the policy objective used to illustrate the nudge, using the same procedure as in our previous studies.

Data exclusions. Participants sometimes completed the survey more than once, so we dropped four responses that indicated a duplicate IP address; including duplicate IP addresses in the analysis does not change the coefficient signs or significance levels of our results. As with studies 1 and 2, we also excluded participants ($n = 4$) who did not report an identifiable social or economic political orientation (that is, they reported being "completely unsure" or that they "haven't given it much thought").

Analysis strategy. We expected that more liberal participants would report greater support for behavioural policy tools when illustrated by liberal objectives rather than conservative objectives, while more conservative participants would support the tools when illustrated by conservative objectives. To test this, we estimated the following ordinary least squares model:

$$\begin{aligned} \text{NudgeEvaluations}_{ij} &= \alpha + \beta_1 \text{Partisan}_i + \beta_2 \text{Application}_j \\ &+ \beta_3 \text{Partisan}_i \times \text{Application}_j + \epsilon_{ij} \end{aligned}$$

where $\text{NudgeEvaluations}_{ij}$ represents attitudes towards automatic enrollment by participant i in treatment condition j . Partisan_i is the self-reported political orientation for participant i and takes on a value between 1 and 7; a positive coefficient indicates that more conservative participants reported greater support for the use of the behavioural policy tool. Application_j is an indicator taking a value of 1 if automatic enrollment was illustrated by a conservative policy goal and a value of 0 if the policy tool was illustrated by a liberal policy goal; a positive coefficient indicates greater support for behavioural policy tools when illustrated by a conservative rather than liberal policy application. The interaction term reflects the change in Application_j as a function of participants' political orientations; a positive interaction term would provide support for partisan nudge bias. We implemented robust standard errors (ϵ_{ij}) to account for arbitrary forms of heteroskedasticity. To provide an estimate of the size for the interaction effect, we performed the same standardized 'spotlight' test as that used in study 1.

The test statistics for conventional (parametric) approaches rely on properties that arise from large samples. Since the sample sizes of studies 3 and 4 were relatively small (especially study 4), we took some additional precautions. We calculated one-tailed *P* values using exact permutation tests, which are independent of sample size. To do this, we conducted 10,000 permutations of our exogenous treatment variable Application_j and calculated the test statistic for the interaction term $\text{Partisan}_i \times \text{Application}_j$; exact *P* values were then calculated by comparing our observed values with these permutations. As a robustness check, we also conducted (large-sample) bootstrapping procedures and found similar results.

Study 4. Participants. We recruited a sample of 48 US city mayors with an average constituency base of 108,000 citizens (range: 8,000–620,000 citizens). Respondents were approached at a US Conference of Mayors event by a representative from the Harvard Kennedy School's Institute of Politics, and were asked to complete a one-page survey in return for a copy of Thaler and Sunstein's book *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Survey administrators were blind to experimental conditions and to the study hypothesis. There was no sample size target; however, we aimed to include as many eligible participants as we could.

Design. The design was identical to study 3 except that the policy nudge was illustrated by automatically enrolling public high school students in supplemental educational programmes for either safe sex practices ($n = 27$) or intelligent design ($n = 21$). Another difference 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 $\alpha = 0.89$) and this time reported their general political orientation using a single-item scale.

Data exclusions and analysis strategy. No data were excluded from study 4. Our analysis strategy was identical to study 3.

Data availability. Data for all studies are available at <https://github.com/davetannenbaum/partisan-nudge-bias>. Included are both raw and 'cleaned' data files.

Code availability. Stata do-files for each study are available at <https://github.com/davetannenbaum/partisan-nudge-bias>. Included are do-files that prepare and clean the raw data files, and analysis scripts that execute all the analyses reported in this paper.

Received 24 December 2016; accepted 23 May 2017;
published 10 July 2017

References

- Madrian, B. C. Applying insights from behavioral economics to policy design. *Ann. Rev. Econom.* **6**, 663–688 (2014).
- Halpern, D. & Sanders, M. Nudging by government: progress, impact, & lessons learned. *Behav. Sci. Policy* **2**, 53–65 (2016).
- Sunstein, C. R. *Simpler: The Future of Government* (Simon and Schuster, 2014).
- Using behavioral science insights to better serve the American people. *Federal Register* **80**, 181 (15 September 2015).
- Congdon, W. J. & Shankar, M. The White House Social & Behavioral Sciences Team: lessons learned from year one. *Behav. Sci. Policy* **1**, 77–86 (2015).
- Allcott, H. Social norms and energy conservation. *J. Public Econ.* **95**, 1082–1095 (2011).
- Madrian, B. C. & Shea, D. F. The power of suggestion: inertia in 401(k) participation and savings behavior. *Q. J. Econ.* **116**, 1149–1187 (2001).
- Thaler, R. H. & Sunstein, C. R. *Nudge: Improving Decisions About Health, Wealth, and Happiness* (Yale Univ. Press, 2008).
- Benartzi, S. *et al.* Should governments invest more in nudging? *Psychol. Sci.* <http://dx.doi.org/10.1177/0956797617702501> (2017).
- Bohannon, J. Government 'nudges' prove their worth. *Science* **352**, 1042–1042 (2016).
- Dunt, I. Nudge nudge, say no more. Brits' minds will be controlled without us knowing it. *The Guardian* (5 February 2014).
- Wyatt, S. Right-wing media characterize government effort to reduce fraud, error, and debt as "mind control". *Media Matters for America* (31 July 2013).
- Page, B. I. & Shapiro, R. Y. Effects of public opinion on policy. *Am. Polit. Sci. Rev.* **77**, 175–190 (1983).
- Hagman, W., Andersson, D., Västfjäll, D. & Tinghög, G. Public views on policies involving nudges. *Rev. Phil. Psych.* **6**, 439–453 (2015).
- Jung, J. Y. *et al.* American attitudes toward nudges. *Judgm. Decis. Mak.* **11**, 62–74 (2016).
- Junghans, A. F., Cheung, T. T. & De Ridder, D. D. Under consumers' scrutiny: an investigation into consumers' attitudes and concerns about nudging in the realm of health behavior. *BMC Public Health* **15**, 336 (2015).
- Sunstein, C. R. Do people like nudges? *Admin. L. Rev.* <http://dx.doi.org/10.2139/ssrn.2604084> (2016).
- Reisch, L. A. & Sunstein, C. R. Do Europeans like nudges? *Judgm. Decis. Mak.* **11**, 310–325 (2016).
- Reisch, L. A., Sunstein, C. R. & Gwozdz, W. Beyond carrots and sticks: Europeans support health nudges. *Food Policy* **69**, 1–10 (2017).
- Sunstein, C. R. *The Ethics of Influence: Government in the Age of Behavioral Science* (Cambridge Univ. Press, 2016).
- Lord, C., Ross, L. & Lepper, M. Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. *J. Pers. Soc. Psychol.* **37**, 2098–2109 (1979).
- Crawford, J. T., Kay, S. A. & Duke, K. E. Speaking out of both sides of their mouths: biased political judgments within (and between) individuals. *Soc. Psychol. Person. Sci.* **6**, 422–430 (2015).
- Kahan, D. M., Jenkins-Smith, H. & Braman, D. Cultural cognition of scientific consensus. *J. Risk Res.* **14**, 147–174 (2011).
- Bolsen, T., Druckman, J. N. & Cook, F. L. The influence of partisan motivated reasoning on public opinion. *Polit. Behav.* **36**, 235–262 (2014).
- Bullock, J. G. Elite influence on public opinion in an informed electorate. *Am. Polit. Sci. Rev.* **105**, 496–515 (2011).
- Druckman, J. N., Peterson, E. & Slothuus, R. How elite partisan polarization affects public opinion formation. *Am. Polit. Sci. Rev.* **107**, 57–79 (2013).
- Taber, C. S. & Lodge, M. Motivated skepticism in the evaluation of political beliefs. *Am. J. Polit. Sci.* **50**, 755–769 (2006).
- Cohen, G. L. Party over policy: the dominating impact of group influence on political beliefs. *J. Pers. Soc. Psychol.* **85**, 808–822 (2003).
- Kahneman, D. Maps of bounded rationality: psychology for behavioral economics. *Am. Econ. Rev.* **93**, 1449–1475 (2003).
- Kahneman, D. & Frederick, S. in *The Cambridge Handbook of Thinking and Reasoning* (eds Holyoak, K. & Morrison, R.) 267–293 (Cambridge Univ. Press, 2005).
- Sah, S., Robertson, C. T. & Baughman, S. B. Blinding prosecutors to defendants' race: a policy proposal to reduce unconscious bias in the criminal justice system. *Behav. Sci. Policy* **1**, 69–76 (2015).
- Kriss, P. H., Loewenstein, G., Wang, X. & Weber, R. A. Behind the veil of ignorance: self-serving bias in climate change negotiations. *Judgm. Decis. Mak.* **6**, 602–615 (2011).
- Bohnet, I., van Geen, A. & Bazerman, M. When performance trumps gender bias: joint vs. separate evaluation. *Manag. Sci.* **62**, 1225–1234 (2015).
- Science and Technology Select Committee *Behaviour Change 2nd Report of Session 2010–12* (House of Lords, 2011); <https://www.publications.parliament.uk/pa/ld201012/ldselect/ldscstech/179/179.pdf>
- Simmons, J. P., Nelson, L. D. & Simonsohn, U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **22**, 1359–1366 (2011).
- Abadie, A. & Gay, S. The impact of presumed consent legislation on cadaveric organ donation: a cross-country study. *J. Health Econ.* **25**, 599–620 (2006).
- Johnson, E. J. & Goldstein, D. G. Do defaults save lives? *Science* **302**, 1338–1339 (2003).
- Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D. & Madrian, B. C. Using implementation intentions prompts to enhance influenza vaccination rates. *Proc. Natl Acad. Sci. USA* **108**, 10415–10420 (2011).
- Nickerson, D. W. & Rogers, T. Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making. *Psychol. Sci.* **21**, 194–199 (2010).
- Rogers, T., Milkman, K. L., John, L. K. & Norton, M. I. Beyond good intentions: prompting people to make plans improves follow-through on important tasks. *Behav. Sci. Policy* **1**, 33–41 (2015).
- Keller, P. A., Harlam, B., Loewenstein, G. & Volpp, K. G. Enhanced active choice: a new method to motivate behavior change. *J. Consum. Psychol.* **21**, 376–383 (2011).
- Fryer, R. G. Jr, Levitt, S. D., List, J. & Sadoff, S. *Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment* Working Paper No. 18237 (National Bureau of Economic Research, 2012).
- Meeker, D. *et al.* Nudging guideline-concordant antibiotic prescribing: a randomized clinical trial. *JAMA Intern. Med.* **174**, 425–431 (2014).
- Rogers, T., Milkman, K. L. & Volpp, K. G. Commitment devices: using initiatives to change behavior. *JAMA* **311**, 2065–2066 (2014).
- Goldstein, N. J. & Cialdini, R. B. in *The Science of Social Influence: Advances and Future Progress* (ed. Pratkanis, A. R.) 167–191 (Psychology Press, 2007).
- Allcott, H. & Rogers, T. The short-run and long-run effects of behavioral interventions: experimental evidence from energy conservation. *Am. Econ. Rev.* **104**, 3003–3037 (2014).
- Spiller, S. A., Fitzsimons, G. J., Lynch, J. G. Jr & McClelland, G. H. Spotlights, floodlights, and the magic number zero: simple effects tests in moderated regression. *J. Market. Res.* **50**, 277–288 (2013).
- Schmidt, F. L. & Hunter, J. E. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings* (Sage Publications, 2014).
- Corey, D. M., Dunlap, W. P. & Burke, M. J. Averaging correlations: expected values and bias in combined Pearson *r*s and Fisher's *z* transformations. *J. Gen. Psychol.* **125**, 245–261 (1998).

Acknowledgements

We are grateful to D. King at the Harvard Kennedy School for assistance with the data collection for study 3, and to C. McLaughlin and C. Flynn at the Harvard Institute of Politics for assistance with the data collection for study 4. We also thank D. Walters and C. Erner for their useful comments on an earlier draft of this paper. No funders had any role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

D.T., C.R.F. and T.R. designed all the experiments, D.T. oversaw the data collection for experiments 1 and 2, T.R. oversaw the data collection for experiments 3 and 4, D.T. analysed the data for all the experiments in consultation with C.R.F., and D.T., C.R.F. and T.R. wrote the manuscript.

Additional information

Supplementary information is available for this paper.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.T.

How to cite this article: Tannenbaum, D., Fox, C. R. & Rogers, T. On the misplaced politics of behavioural policy interventions. *Nat. Hum. Behav.* **1**, 0130 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Competing interests

The authors declare no competing interests.