Behavioural Public Policy (2020), 4: 3, 273–291
© Cambridge University Press doi:10.1017/bpp.2018.9

Are 'nudges' getting a fair shot? Joint versus separate evaluation

SHAI DAVIDAI*

The New School for Social Research, New York, NY, USA

ELDAR SHAFIR

Princeton University, Princeton, NJ, USA

Abstract: The most effective behavioral policies are often also the most contentious. Psychologically informed interventions that promote non-deliberative behaviors ('nudges') are often more effective than 'traditional' policies (like informational and educational campaigns) that target more deliberative processes. Yet, precisely because of their deliberative nature, people are often said to prefer the latter over the former. In contrast, we provide evidence that people's preferences regarding nudges are malleable and influenced by the method of evaluation – whether the policy alternatives are evaluated separately or jointly. We show that while people exhibit a strong preference for more traditional public policies in joint evaluation, this preference is significantly attenuated in separate evaluation. We find that people perceive nudges as less paternalistic when judged on their own merits, that they are more likely to endorse nudges in separate than in joint evaluation, and that, provided with relative effectiveness information, people are willing to endorse nudges even in joint evaluation. We discuss the implications of these findings for researchers, policy-makers, and the general public.

Submitted 26 February 2018; accepted 8 March 2018

What if we could make people healthier, wealthier, and happier, at a relatively low cost, and without infringing on their freedom? In the past decade or so, psychologically informed policies, and especially 'nudges', have been trying to do just that (Thaler & Sunstein, 2008). Insights from psychology and behavioral economics have shifted the way policy-makers approach social issues such as health, wealth, nutrition, education, voter turnout, judicial decision-making,

* Correspondence to: The New School for Social Research, 80 5th Ave, Room 708, New York, NY 10011, USA. Email: shaidavidai@newschool.edu

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

and environmentally sustainable behavior, to name just a few (Benartzi & Thaler, 1999; Johnson & Goldstein, 2003; Wansink, 2006; Goldstein *et al.*, 2008; Nickerson & Rogers, 2010; Bryan *et al.*, 2011; Danziger *et al.*, 2011; Dayan & Bar-Hillel, 2011; Milkman *et al.*, 2011; Rozin *et al.*, 2011; Walton & Cohen, 2011; Davidai *et al.*, 2012; Allcott & Rogers, 2014; Thorndike *et al.*, 2014). The evidence suggests that this shift toward behavioral policies has been quite effective. A recent comprehensive analysis has found the impact-to-cost ratio of various nudges (e.g., default vaccination appointments) to be significantly higher than the impact-to-cost ratio of more traditional policies (e.g., monetary incentives for vaccinations; Benartzi *et al.*, 2017). For example, nudging people to conserve energy by activating social norms resulted in almost twice the reduction in kWh for every $1 spent compared to a traditional educational campaign around energy-efficient consumption. Similarly, retirement saving rates more than double when employees are enrolled into a savings program that automatically escalates their contribution with every salary raise (Thaler & Benartzi, 2004). Impressive results of this kind have gained psychologists a seat at a table traditionally reserved for economists and policy advisors, as exemplified by President Barack Obama's executive order to "realize the benefits of behavioral insights and deliver better results at a lower cost for the American people" (Office of the Press Secretary, 2015).

Yet not all is quiet on the 'nudge' front. Policy-makers "are inevitably responsive to what people think" (Reisch & Sunstein, 2016, p. 311), and whether a potential nudge becomes reality hinges not only on its expected benefit to society, but also on its ability to garner public and legislator support. Unpopular nudges can backfire, leading people to react against what are seen as illicit attempts to shape their behavior (Arad & Rubinstein, 2018; Damgaard & Gravert, 2017; Jachimowicz *et al.*, 2018), undermine their autonomy (Dunt, 2014; Shaw, 2017), and compromise their dignity "in the way one indulges a child" (Waldron, 2014). Worse still, negative reactions toward specific 'nudges' can lead people to question the entire practice of using behavioral research in policy (Tannenbaum *et al.*, 2017). For nudges to be implemented, it is imperative that they not encounter strong public resistance.

How, then, do people feel about nudges? Although highly cost-effective interventions often target automatic, non-deliberative ('System 1') decision processes, people are generally more suspicious of these than they are of ones that target deliberate and conscious ('System 2') decision processes.[1]

1 In keeping with previous research (e.g., Jung & Mellers, 2016; Sunstein, 2016a, 2016b), and while clearly an oversimplification, we occasionally use the terms 'System 1' and 'System 2' nudges as *placeholders* for the array of policies that target non-deliberative and deliberative psychological processes, respectively (see Kahneman, 2011).

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

Motivated by a concern for their autonomy and freedom of choice, people prefer policies that target processes of which they are fully aware (e.g., educational and public awareness campaigns, financial literacy programs, etc.) over those that target passive or peripheral processes that often go unnoticed or unappreciated (e.g., automatic enrollment, default options, emotionally vivid graphics, etc.; Jung & Mellers, 2016; Reisch & Sunstein, 2016; Sunstein, 2016b). In this vein, it has been argued that people prefer policies that provide information about the best course of action (e.g., nutritional value of food options) over those (like the design of menus or cafeterias) that 'nudge' people into taking the healthier options (Felsen *et al.*, 2013; Arad & Rubinstein, 2018).

Clearly, people are not *categorically* opposed to nudges, and some nudges are quite popular. Whereas people prefer policies that target deliberative processes over nudges, Sunstein (2016b) reports that, given the option, many people support policies that target both deliberative *and* non-deliberative processes over ones that are exclusively 'System 2'.[2]

Because public opinion can shape policy-makers' decisions, it is important to understand whether surveys of public opinion indeed reflect the public's views. Do people really have deep-rooted preferences for deliberative policies over non-deliberative nudges, or are their preferences malleable, mostly a function of how people are surveyed? Are we giving nudges 'a fair shot'?

In this paper, we argue that attitudes toward nudges are not as stable, nor as negative, as often portrayed. We suggest that previous research has inadvertently exaggerated the preference for deliberate policy interventions over ones that target non-deliberative processes (e.g., Jung & Mellers, 2016; Sunstein, 2016b). Following a long tradition in judgment and decision-making research, we argue that the apparent preference for deliberative, 'System 2' policies may be largely an artifact of the way options have been evaluated. In particular, we argue that the direct contrast between the two types of intervention renders prominent the defining difference between deliberative and non-deliberative policies in ways that independent evaluation fails to do (Hsee *et al.*, 1999).

An extensive body of research has shown that different mental processes are often engaged when two options are simultaneously presented and can be directly compared (i.e., joint evaluation) as opposed to when options are presented separately, each evaluated in isolation (i.e., separate evaluation;

2 Preference for policies that target both deliberative *and* non-deliberative processes over ones that target purely non-deliberative processes (Sunstein, 2016b) further supports the notion that people are relatively uneasy with pure nudges. Policy-makers can thus increase a nudge's popularity by adding components of more traditional, 'System 2' policies to it.

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

Bazerman *et al.*, 1999; Hsee *et al.*, 1999). Sometimes, when numeric values are hard to gauge, joint evaluation facilitates comparison (Hsee, 1996, 1998). In other cases, a qualitative dimension, which plays a minor role in separate evaluation, becomes more salient in direct comparison.

In one study, for example, Kahneman and Ritov (1994) presented people with two proposed interventions, one concerning human health checkups and the other concerning the protection of an endangered animal species. When the interventions were evaluated separately, respondents expressed a willingness to pay more for the protection of the endangered animals than for people's medical checkups. But when faced with a direct choice between the two, most respondents favored free checkups for humans over protection for animals. As it turns out, the 'humans come before (non-human) animals' consideration played a critical role in direct comparison, but had much less force in separate evaluation.

Interestingly, while the 'humans first' argument comes naturally to mind when contributions to humans and contributions to other animals are jointly evaluated, it is not clear that it captures people's 'true sentiments'. After all, substantial amounts of money are donated each year to many animal protection causes, with US$3 billion spent by animal shelters, US$1.7 billion spent on the Endangered Species Act, and over US$60 billion spent on pets in the USA alone. When people contemplate costly interventions in the absence of a direct human–animal contrast (as in the studies above), some human-related causes lose their priority over those concerning the protection of animals (see also Irwin *et al.*, 1993; Kahneman *et al.*, 1999).

Similar to the pattern above, we argue that when people directly compare between deliberative and non-deliberative interventions, they are overly influenced by the contrast made salient by this comparison, which can become rather minor and largely inconsequential in separate evaluation, when policies' actual merits loom larger. In what follows, we argue that joint evaluation highlights concerns about deliberation, freedom of choice, and paternalism and, as a consequence, reduces support for non-deliberative nudges (Mellers & Jung, 2016). Those same concerns, however, become rather negligible in people's minds when evaluating each policy based on its own merit. 'Libertarian paternalistic' nudges (Thaler & Sunstein, 2008), in other words, appear significantly less libertarian and more paternalistic in joint evaluation. A nudge that seems paternalistic in joint evaluation may not seem so when evaluated in isolation.

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

Studies 1A and 1B: how do people really feel about nudges?

Method

Participants

We recruited participants from Amazon's Mechanical Turk in exchange for a modest payment. As part of two longer experimental sessions, 113 people participated in Study 1A (70 females, $M_{age} = 36.06$) and 130 participated in Study 1B (62 females, $M_{age} = 35.42$).

Manipulation and dependent variables

We conducted two studies, adapting materials from Jung and Mellers (2016), to examine whether the mode of evaluation (joint vs. separate evaluation) influences people's preference for policies that target deliberative ('System 2') over non-deliberative ('System 1') processes. In Study 1A, we asked participants whether they would support a new policy designed to increase enrollment in a basic medical insurance plan. We varied whether the policy involved automatic (non-deliberative, 'System 1' nudge) employee enrollment or an educational (deliberative, 'System 2') campaign about the plan's benefits:

> *Suppose that when you start a new job at a company, the government requires the company to [select a default policy that makes it easier to enroll in a basic medical plan. You have other options, but you need to select different options to implement them. The plan will help you enjoy the benefits of good health over the long term.] / [provide information that makes it easier to understand the benefits of enrolling in a basic medical plan. The information encourages you to enroll by explaining why the plan is designed to help you enjoy the benefits of good health over the long term.]*

In Study 1B, we asked participants whether they would support a new policy designed to encourage timely payments of credit card bills. We varied whether the policy involved a full-balance payment as the default, non-deliberative ('System 1') option or, instead, provided information to customers explaining the benefits of making such deliberative ('System 2') payments:

> *Suppose that when you pay your credit card bills online, the government requires the credit card company to [set a default payment option that makes it easy to pay the bill in full. You can pay other amounts, but you need to specify those by selecting different options. The default policy is designed to help you enjoy the benefits of no interest fees and the advantages of excellent credit scores.] / [provide information that makes it easy to understand the benefits of paying the total amount due. The information encourages you to pay the entire bill by telling you that full payment helps*

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

you enjoy the benefits of no interest fees and the advantages of excellent credit scores.]

In each study, we randomly assigned participants to one of three conditions: two in which they evaluated either (and only) one of the policy options (*separate evaluation* conditions); and one in which they evaluated the two policies jointly (*joint evaluation* condition). Participants indicated their support for each policy on a scale ranging from –3 (certainly no) to +3 (certainly yes), with the midpoint (0) indicating a neutral attitude toward the policy.

Results

We first report the results from the *joint evaluation* conditions. In a direct replication of Jung and Mellers (2016), we found that participants preferred policies that targeted deliberative ('System 2') processes over those that targeted non-deliberative ('System 1') processes. Participants in Study 1A were significantly more supportive of a policy that educates about the benefits of a medical plan (M = 1.44, SD = 1.63) than a policy that automatically enrolls employees into such a plan (M = 0.28, SD = 1.77), matched pairs $t(53) = 3.29$, $p = 0.002$. Participants in Study 1B were significantly more supportive of a policy that informs people of the benefits of paying their credit card bill in full (M = 1.84, SD = 1.50) than a policy that sets full-balance payments as the default (M = –0.22, SD = 2.01), matched pairs $t(36) = 5.15$, $p < 0.0001$. Replicating past findings, we found that a direct comparison of the options led participants to express significantly greater support for deliberative policies over non-deliberative nudges.

Now, what about separate evaluation? Were participants' attitudes influenced by the mode of evaluation? As can be seen in Figures 1A and 1B, we observed a clear shift. Participants in Study 1A clearly preferred the deliberative option in joint evaluation, but they were no more likely to support it (M = 1.62, SD = 1.13) over the non-deliberative nudge (M = 1.30, SD = 1.45) when these were evaluated separately, $t(57) < 1$, *ns*. The difference in support of the two policies in joint and separate evaluation was significant, $t(112) = 5.55$, $p < 0.0001$.[3] A similar but less pronounced pattern was observed in Study 1B. While separate evaluation in Study 1B still showed a preference for the deliberative policy option

3 Because the difference in attitude in direct comparison is within participants and the difference in attitude in separate evaluation is between participants, we use the following *t* statistic to compare these differences: $([M_{j1} - M_{j2}] - [M_{s1} - M_{s2}])/([S_j^2/N_j + S_{s1}^2/N_{s1} + S_{s2}^2/N_{s2}^2]^2)$, where M_{j1}, M_{j2}, M_{s1}, and M_{s2} are means for attitudes regarding nudges and traditional policies in direct comparison and separate evaluations, respectively; S_j^2, S_{s1}^2, and S_{s2}^2 are variances; and N_j, N_{s1}, and N_{s2} are the numbers of participants in the direct comparison and the two separate evaluation conditions. See Footnote 2 in Hsee (1996).

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

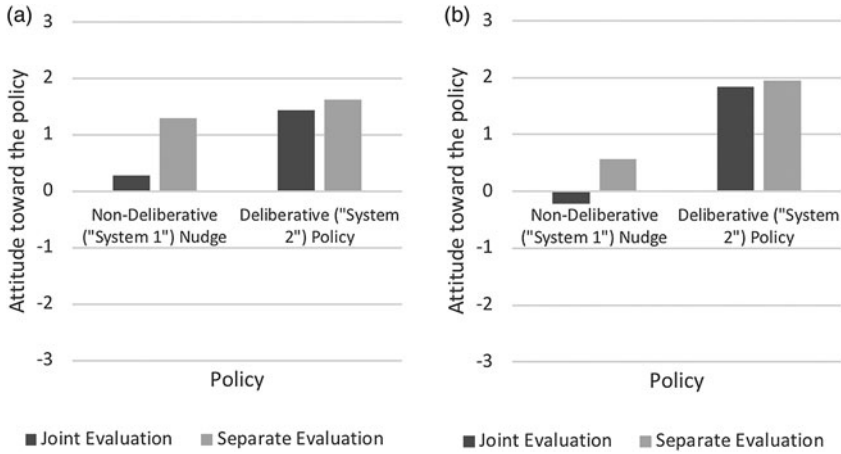


Figure 1. Participants' support for 'System 1' and 'System 2' policies in joint and separate evaluation for policies involving basic medical coverage (A) and credit card bills (B).

($M_{\text{'System 2' policy}} = 1.95$, SD = 1.19; $M_{\text{'System 1' nudge}} = 0.56$, SD = 1.99), $t(91) = 4.01$, $p = 0.0001$, this preference was significantly attenuated relative to the joint evaluation condition, $t(129) = 4.78$, $p < 0.0001$.

What accounts for participants' changes in preference? A closer look at the data reveals that in both studies, the mode of evaluation (joint vs. separate) diminished participants' support for the non-deliberative, 'System 1' option, but had no impact on the deliberative, 'System 2' alternative. In Study 1A, participants judged the (deliberative) education policy to be equally favorable both in joint (M = 1.44) and separate (M = 1.62) evaluation, $t(78) < 0.5$, *ns*. In contrast, they were significantly more supportive of the non-deliberative automatic enrollment of employees in the separate (M = 1.30) than the joint (M = 0.28) evaluation, $t(85) = 2.80$, $p = 0.006$. Similarly, participants judged the deliberative policy in Study 1B equally favorably in joint (M = 1.84) and in separate (M = 1.95) evaluation, $t(78) < 0.5$, *ns*, but were marginally (albeit not statistically significantly) more supportive of the non-deliberative nudge in the separate (M = 0.56) rather than the joint (M = –0.22) evaluation, $t(85) = 1.78$, $p = 0.078$. Whereas the appeal of deliberative policies was unaltered, support for non-deliberative policy options diminished once the deliberative contrast between them was made salient.

Study 2: can nudges garner majority support?

We found that participants' attitudes towards non-deliberative policies were affected by the mode of evaluation. Whereas deliberative policies were

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

preferred over non-deliberative nudges in joint evaluation, this preference was lost or significantly attenuated in separate evaluation. Non-deliberative, 'System 1' nudges that were judged quite appealing when evaluated on their own merits lost some of that appeal when in direct comparison with seemingly 'less manipulative' policies targeting deliberative processes. In Study 2, instead of rating their attitude, we asked participants to indicate support or opposition to the policies.

Method

Participants

Three hundred participants from Amazon's Mechanical Turk took part in the study (230 females, $M_{age} = 35.05$).

Manipulation and dependent variables

Participants read a short paragraph describing a new policy designed to encourage people to save more toward retirement (see Jung & Mellers, 2016). We varied whether the policy involved a requirement to automatically enroll employees into a basic retirement savings plan (non-deliberative nudge) or a requirement to inform employees of the benefits of enrolling in such a plan (traditional deliberative policy):

> *Suppose that when you start a new job at a company, the government requires the company to [select a default policy that makes it easy to enroll in a basic retirement savings plan. You have other options, but you need to select those to indicate a different set of preferences. The default policy of basic retirement savings is designed to help you enjoy a comfortable lifestyle and feelings of financial security when you stop working.] / [provide information that makes it easy to understand the benefits of enrolling in a basic retirement savings plan. The information encourages you to enroll in the basic savings plan by telling you that the plan is designed to help you enjoy a comfortable lifestyle and feelings of financial security when you stop working.]*

Participants were randomly assigned to one of three conditions: two conditions in which they evaluated either (and only) one of the two policies (*separate evaluation* conditions); and one in which they evaluated both policies at the same time (*joint evaluation* condition). Participants were asked whether they supported, opposed, or had no opinion about the policies.

Results

As in Studies 1A and 1B, we found that the mode of evaluation (joint vs. separate) significantly influenced policy support. In the joint evaluation condition,

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

89% of participants supported the deliberative ('System 2') policy informing employees of the benefits of enrolling in a retirement savings plan, whereas only 43% supported the non-deliberative ('System 1') nudge automatically enrolling employees in such a plan, $\chi^2(1) = 47.46$, $p < 0.0001$. This difference, however, was significantly attenuated in the separate evaluation conditions, where 74% of participants supported the deliberative policy and 61% supported the non-deliberative nudge, $\chi^2(1) = 3.98$, $p = 0.046$. As before, participants were significantly more supportive of the nudge when evaluating it based on its own merits than when comparing it to the seemingly 'less intrusive' alternative. Whereas support for the non-deliberative option increased from 43% in joint evaluation to 61% in separate evaluation, $\chi^2(1) = 6.22$, $p = 0.013$, support for the deliberative policy actually diminished from 89% in the joint evaluation to 74% in the separate evaluation condition, $\chi^2(1) = 7.21$, $p = 0.007$.

It is notable that both types of policy garnered majority support in all but one case. The majority of participants supported the deliberative policy both in joint and in separate evaluation, χ^2s > 25.00, ps < 0.0001, and supported the non-deliberative nudge in the separate evaluation condition, $\chi^2(1) = 4.88$, $p = 0.027$. Only in the joint evaluation condition, when compared to the deliberative alternative, did the non-deliberative nudge fail to garner majority support, $\chi^2(1) = 1.75$, *ns*. The same nudge that garnered majority support when evaluated on its own would have failed to do so in joint evaluation.

Study 3: are nudges paternalistic?

When people evaluate policies separately, their preference for deliberative policies over non-deliberative nudges is significantly attenuated, if not fully eliminated. In joint evaluation, in contrast, the deliberative versus non-deliberative nature of policies outweighs their independent merit and renders non-deliberative nudges less appealing. In the present study, we look a bit deeper into the potential causes of this shift. Specifically, we examine whether the salient tension between policies' deliberative versus non-deliberative nature leads participants to perceive nudges as more paternalistic in joint than in separate evaluation. Because concerns about paternalism underlie people's opposition to nudges (Jung & Mellers, 2016) and because the non-deliberative nature of nudges looms larger in direct comparison to deliberative alternatives, we hypothesized that nudges will be perceived as more paternalistic and more restrictive of people's autonomy when directly compared to more deliberative policies than when evaluated in isolation.

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

Method

Participants

One hundred and twenty-three participants from Amazon's Mechanical Turk took part in the study (68 females, $M_{age} = 40.48$).

Manipulation and dependent variables

Participants read the same short paragraph as in Study 2, describing a new policy designed to encourage people to save more toward retirement. We varied whether the policy involved automatic enrollment into a basic retirement savings plan (non-deliberative nudge) or information about the benefits of enrolling in such a plan (traditional deliberative policy). Participants were randomly assigned to one of three conditions: two in which they evaluated either (and only) one of the policies (*separate evaluation* conditions); and one in which they evaluated both policies jointly (*joint evaluation* condition).

Next, participants read a short paragraph describing the difference between libertarian and paternalistic policies:

> *Initiatives differ in how paternalistic or libertarian they seem. On one end of the spectrum are extremely paternalistic initiatives which completely restrict people's ability to decide for themselves and restrict their freedom of choice. For example, drivers are often required by law to drive below a specified speed limit and don't have a choice of whether they want to do so or not. On the other end of the spectrum are extremely libertarian initiatives which maximize people's freedom and autonomy. Some countries, for example, don't have speed limits and drivers can freely choose how fast they wish to drive.*

Participants then indicated the extent to which they deemed the proposed policy paternalistic versus libertarian on a six-point Likert scale (1 – Extremely paternalistic, 2 – Moderately paternalistic, 3 – A little paternalistic, 4 – A little libertarian, 5 – Moderately libertarian, 6 – Extremely libertarian).

Results

As predicted, the extent to which the nudge was deemed paternalistic depended on the method of evaluation (Figure 2). In the joint evaluation condition, participants perceived the non-deliberative nudge to automatically enroll employees into savings as significantly more paternalistic ($M = 2.95$, $SD = 1.38$) than a deliberative policy that informed employees about the benefits of enrollment ($M = 3.92$, $SD = 1.38$), matched pairs $t(38) = 2.45$, $p = 0.019$. In contrast, participants in the separate evaluation conditions – who judged each policy based on its own merits – perceived the two policies as equally paternalistic

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

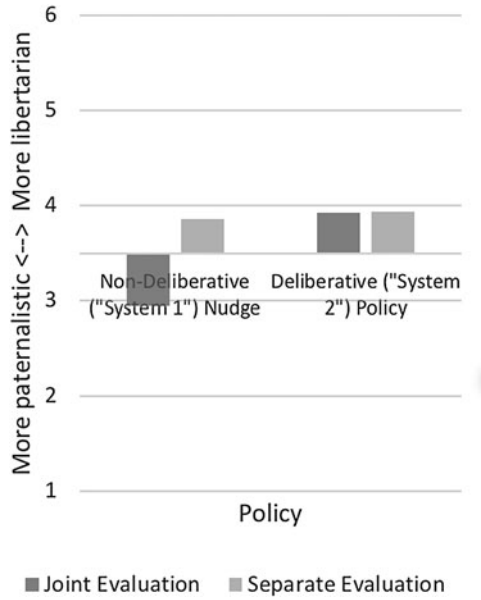


Figure 2. Participants' judgments of paternalism of 'System 1' and 'System 2' policies in joint and separate evaluation (Study 3).

($M_{\text{'System 1' nudge}} = 3.86$, SD = 1.37; $M_{\text{'System 2' policy}} = 3.93$, SD = 1.24), $t(82) < 0.5$, *ns*. The difference in paternalism judgment of the two policies in joint and separate evaluation was significant, $t(122) = 13.66$, $p < 0.0001$.

As before, perceptions of the nudge (but not of the traditional policy) were significantly influenced by the method of evaluation. Whereas the deliberative policy was deemed equally paternalistic in joint (M = 3.92) and in separate evaluation (M = 3.93), $t(79) < 0.1$, *ns*, the non-deliberative nudge was seen as significantly more paternalistic in joint (M = 2.95) than in separate evaluation (M = 3.86), $t(79) = 2.97$, $p < 0.005$.

Study 4: can nudges garner support in joint evaluation?

In Study 4, we examined whether explicit information about initiatives' projected merit, or effectiveness, can garner support for non-deliberative ('System 1') policies even in joint evaluation, when they are typically less likely to be favored (see also Sunstein, 2016b). Earlier work on joint versus separate evaluation has shown that numerical information, which is sometimes hard to gauge in separate evaluation, figures more prominently in joint evaluation, where the advantages of one option over another become transparent

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

(Hsee, 1996). In this vein, because efficiency is hard to evaluate in isolation without an explicit benchmark, we expected that effectiveness information would have little influence on participants' attitudes in separate evaluation. In contrast, because attitudes toward nudges appear rather malleable, we predicted that when directly comparing between policies whose merits are explicit, participants would prefer more effective policies regardless of whether they target deliberative or non-deliberative processes.

Methods

Participants

One hundred participants from Amazon's Mechanical Turk took part in the study (61 females, $M_{age} = 34.82$).

Manipulation and dependent variables

Participants read the same short paragraph as in Study 1A, describing a new policy designed to encourage enrollment in a basic medical insurance plan. We varied whether the policy involved automatic enrollment of employees into a basic medical plan (non-deliberative nudge) or an educational campaign about such a plan (deliberative policy). While both policies were projected to increase insurance coverage, the non-deliberative enrollment nudge was shown to be very effective ("projected to increase medical coverage by more than 40%"), whereas the deliberative information intervention was somewhat effective ("projected to increase medical coverage by almost 5%").[4] As before, participants were randomly assigned to one of three conditions: two in which they evaluated one of the two policies (*separate evaluation* conditions); and one in which they evaluated both policies jointly (*joint evaluation* condition). Participants indicated whether they would support each policy on a sliding scale ranging from –3 (certainly no) to +3 (certainly yes).

Results

As shown in Figure 3, the proposed policies' effectiveness information dominated participants' judgments when the policies were evaluated jointly, but not when they were judged separately. When participants evaluated the policies separately, they were equally favorable toward both plans ($M_{\text{'System 1' nudge}} = 1.13$, $M_{\text{'System 2' policy}} = 1.55$), $t(74) < 1.4$, *ns*. In contrast, when they evaluated

4 Because the scenario involved two hypothetical policies, we chose values that reflect the general relative effectiveness of default policies. As discussed in Benartzi and colleagues (2017), such large differences in effectiveness are not uncommon. For a similar approach, see Sunstein (2016b).

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

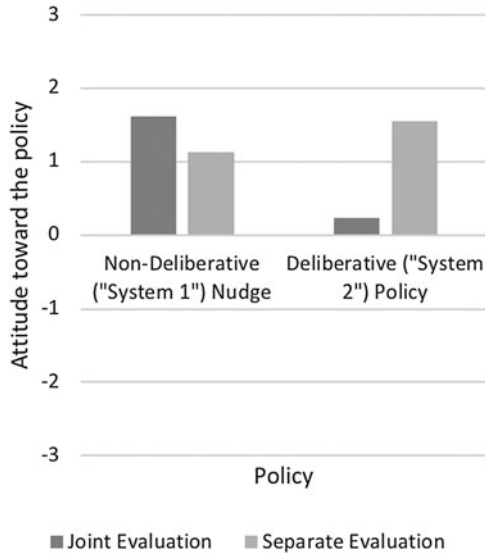


Figure 3. Participants' support for the more effective nudge and the less effective traditional medical coverage policy in joint and separate evaluation (Study 4).

the policies jointly, participants were significantly more supportive of the highly effective non-deliberative nudge (M = 1.61, SD = 1.24) than the less effective but deliberative policy (M = 0.23, SD = 1.58), $t(25) = 3.25$, $p = 0.003$. The difference in expressed support for the two policies in joint and in separate evaluation was significant, $t(99) = 11.64$, $p < 0.001$. Because participants could easily compare the effectiveness in the joint evaluation, they were significantly more supportive of the more effective yet non-deliberative nudge than of its deliberative but less effective alternative.

General discussion

How do people really feel about 'nudges'? Although past research has argued that people have a clear-cut preference for policies that target deliberative ('System 2') over non-deliberative ('System 1') processes (Felsen *et al.*, 2013; Jung & Mellers, 2016; Sunstein, 2016a, 2016b), we find that things are more subtle. Attitudes are malleable and can be influenced by the method of evaluation – in this case, whether policies are evaluated separately or jointly. Joint evaluation tends to highlight what is salient in direct comparison, which, in the context of the present studies, was the more or less deliberative

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

nature of policy recommendations. Indeed, we have found that non-deliberative nudges are seen as more paternalistic and more objectionable than deliberative alternatives when these are evaluated jointly, but that this is reduced or fully eliminated when they are evaluated separately.

Given the differences produced by mode of evaluation, which mode should be used to give nudges 'a fair shot'? Or alternatively, should researchers (and policy-makers) be more concerned with attitudes expressed in joint or in separate evaluation? One could argue that joint evaluation is preferable precisely because it highlights the deliberative versus non-deliberative aspects of policies. Yet, normatively speaking, it is not clear why this highlight should prove particularly beneficial. As observed by Hsee and Zhang (2004), considerations rendered salient in joint evaluation are not always the most normatively compelling. Because joint evaluation highlights certain aspects largely independent of their normative appeal, there is unlikely to be a simple solution to which mode of presentation, joint or separate, proves to be normatively superior.

It is worth noting that policies are typically implemented in one version or another; rarely do people get to entertain multiple policy options simultaneously. Whereas policy-makers and academics may be in a unique position of 'jointly evaluating' alternative interventions, the general public is more often than not confronted with a specific policy, effectively engaging in 'separate evaluation' (Shafir, 1999; Hsee & Zhang, 2004). And because policies are ultimately experienced in isolation, attitudes elicited in joint evaluation may be less reflective of people's real experiences. Policy-makers are often in the position to directly compare several potential policies with one another, weighing their pros and cons. The general public, in contrast, is rarely in that same privileged position. People experience and react to policies in what amounts to a separate evaluation, whereas policy-makers have the uncommon experience – and responsibility – of evaluating policies comparatively.

At times, people may engage in what resembles a joint evaluation, particularly when comparing an experience to possible alternatives. For example, when contemplating an automatic enrollment policy, people may spontaneously bring to mind an alternative, such as the one about to be replaced or one in use at another institution or location. Judgments can be influenced not only by focal events, but also by counterfactuals that naturally spring to mind (Kahneman & Miller, 1986). This could explain why preference for deliberative policies in our studies was significantly attenuated in separate evaluation, but not always fully eliminated. When we asked participants to evaluate fairly novel and non-deliberative interventions (e.g., default total-balance payments of bills), more familiar and less 'intrusive' policies may have been automatically generated. As behaviorally informed interventions grow in popularity and the public becomes more familiar with non-deliberative

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

nudges, people may be less prone to compare them with salient deliberative alternatives, and may thus become more supportive.

Of course, nudges are not a policy-maker's panacea, and social problems will often need to be addressed with other types of policy interventions. At times, governments may impose highly paternalistic policies, as with rules prohibiting underage alcohol consumption. Other times, they resort to less paternalistic policies such as educational campaigns against underage drinking.[5] Given the growing interest in behavioral interventions, we have chosen to focus on people's attitudes toward nudges – policies that are neither fully paternalistic, nor completely libertarian (Thaler & Sunstein, 2008). Because such policies often involve an implicit trade-off between effectiveness and the concern with 'being nudged', people's attitudes toward them are prone to be influenced by the mode of evaluation. Yet, in cases when this trade-off is one-sided (e.g., if a nudge is seen as very ineffective or extremely offensive), people may be inclined to oppose the policy no matter how evaluation is elicited. For example, Sunstein (2016b) observes that some nudges (e.g., requiring women to see vivid *in utero* photos before deciding to abort a pregnancy) fail to garner majority support regardless of their effectiveness. Of course, the more pronounced people's attitudes toward a particular nudge, the less we would expect them to be affected by mode of evaluation.

Because mode of evaluation is likely to prove most consequential in cases where preferences are less pronounced, it is particularly important to understand the subtle influence of evaluation mode on matters pertinent to policy-making. Doing so could guide researchers in choosing the modes that most closely tap into how laypeople spontaneously think about their decisions in everyday life (Davidai & Gilovich, 2018), increase the ecological validity of surveyed attitudes about public policies, and form a broader and more comprehensive understanding of how people feel about nudges.

While we have focused here on deliberative versus non-deliberative processes, public policies differ on many dimensions that may be similarly influenced by joint versus separate evaluation. As shown in Study 4, policies' effectiveness is one such dimension. For another, Hagman and colleagues (2015) have found that people tend to support policies that benefit individual over societal well-being, whereas Cornwell and Krantz (2014) have shown

5 Hertwig and Grüne-Yanoff (2017) offer yet another distinction between nudges (policies that target automatic, non-deliberative processes) and 'boosts' (policies that improve individuals' abilities to make informed and competent choices; see also Grüne-Yanoff & Hertwig, 2016). While not the focus of the current work, it is interesting to consider the influence of direct comparisons on this kind of policies. We suspect that, similar to nudges, attitudes toward 'boost' policies (and their perceived paternalism) would be equally influenced by evaluation mode.

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

that, in separate evaluation, people are less supportive of policies that target themselves than those that target people in general. Future research can examine the influence of joint and separate evaluation on these and related findings, including, importantly, ones that involve moral judgments. Thus, while Van Zant and Moore (2015) found that moral as opposed to pragmatic considerations influenced people's support of proposed policies, it is possible that pragmatic considerations, having to do with a policy's actual impact, may play a larger role in joint compared to separate evaluation.

Finally, although it is a widely applicable concern, it is important to keep in mind that mode of evaluation is not the only way in which people's attitudes toward public policies can be shaped. Attitudes toward redistributive policies, for example, are influenced by whether economic inequality is framed as the rich having more or the poor having less (Chow & Galak, 2012); people's perception of economic mobility is affected by whether they consider upward or downward social mobility (Davidai & Gilovich, 2015); and their support for behaviorally informed policies is significantly influenced by who they believe put the policies in place (i.e., the 'partisan nudge bias'; Tannenbaum *et al.*, 2017). These and other factors may influence attitudes in ways that interact with modes of evaluation in subtle and interesting ways. Do people spontaneously generate direct comparisons to more attractive options when faced with proposals that are less appealing to their political taste? Are politicians more likely to introduce 'joint evaluation' when attacking their opponents' proposed (especially non-deliberative) policy plans?

Recent research has examined how policy-makers might increase public support for non-deliberative interventions, or 'nudges' (Loewenstein *et al.*, 2015; Steffel *et al.*, 2016). In this paper, we discussed a simple yet non-intuitive method of increasing such support. Direct comparison between more and less deliberative policies raises the salience of deliberation and reduces support for non-deliberative interventions. But as long as they evaluate policy proposals separately and on their own merit, people seem quite willing to support what are often highly effective 'System 1' nudges. Assuming that what policy-makers ultimately care about are the consequences of policies, not the manner in which they are evaluated, this nuanced distinction is worth paying attention to.

References

Allcott, H. and T. Rogers (2014), 'The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation', *The American Economic Review*, **104** (10): 3003–3037.

Arad, A. and A. Rubinstein (2018), The People's Perspective on Libertarian-Paternalistic Policies. *Unpublished manuscript.*

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

Bazerman, M. H., D. A. Moore, A. E. Tenbrunsel, K. A. Wade-Benzoni and S. Blount (1999), 'Explaining how preferences change across joint versus separate evaluation', *Journal of Economic Behavior & Organization*, **39**(1): 41–58.

Benartzi, S., J. Beshears, K. L. Milkman, C. R. Sunstein, R. H. Thaler, M. Shankar, W. Tucker-Ray, W. J. Congdon and S. Galing (2017), 'Should governments invest more in nudging?' *Psychological Science*, DOI: 10.1177/0956797617702501.

Benartzi, S. and R. H. Thaler (1999), 'Risk aversion or myopia? Choices in repeated gambles and retirement investments', *Management Science*, **45**(3): 364–381.

Bryan, C. J., G. M. Walton, T. Rogers and C. S. Dweck (2011), 'Motivating voter turnout by invoking the self', *Proceedings of the National Academy of Sciences*, **108**(31): 12653–12656.

Chow, R. M. and J. Galak (2012), 'The effect of inequality frames on support for redistributive tax policies', *Psychological Science*, **23**(12): 1467–1469.

Cornwell, J. F. and D. H. Krantz (2014), 'Public policy for thee, but not for me: Varying the grammatical person of public policy justifications influences their support', *Judgment and Decision Making*, **9**(5): 433–444.

Damgaard, M. T. and C. Gravert (2017), 'Now or never! The effect of deadlines on charitable giving: Evidence from two natural field experiments', *Journal of Behavioral and Experimental Economics*, **66**, 78–87.

Danziger, S., J. Levav and L. Avnaim-Pesso (2011), 'Extraneous factors in judicial decisions', *Proceedings of the National Academy of Sciences*, **108**(17): 6889–6892.

Davidai, S. and T. Gilovich (2018), How should we think about Americans' beliefs about economic mobility? Manuscript accepted for publication.

Davidai, S. and T. Gilovich (2015), 'Building a more mobile America—One income quintile at a time', *Perspectives on Psychological Science*, **10**(1): 60–71.

Davidai, S., T. Gilovich and L. D. Ross (2012), 'The meaning of default options for potential organ donors', *Proceedings of the National Academy of Sciences*, **109**(38): 15201–15205.

Dayan, E. and M. Bar-Hillel (2011), 'Nudge to nobesity II: Menu positions influence food orders', *Judgment and Decision Making*, **6**(4): 333–342.

Dunt, I. (2014), Nudge nudge, say no more: Brits' minds will be controlled without us knowing it. *The Guardian*. Retrieved from: https://www.theguardian.com/commentisfree/2014/feb/05/nudge-say-no-more-behavioural-insights-team

Felsen, G., N. Castelo and P. B. Reiner (2013), 'Decisional enhancement and autonomy: public attitudes towards overt and covert nudges', *Judgment and Decision Making*, **8**(3): 202–213.

Grüne-Yanoff, T. and R. Hertwig (2016), 'Nudge versus boost: How coherent are policy and theory?', *Minds and Machines*, **26**(1–2): 149–183.

Goldstein, N. J., R. B. Cialdini and V. Griskevicius (2008), 'A room with a viewpoint: Using social norms to motivate environmental conservation in hotels', *Journal of Consumer Research*, **35**(3): 472–482.

Hagman, W., D. Andersson, D. Västfjäll and G. Tinghög (2015), 'Public views on policies involving nudges', *Review of Philosophy and Psychology*, **6**(3): 439–453.

Hertwig, R. and T. Grüne-Yanoff (2017), 'Nudging and Boosting: Steering or Empowering Good Decisions', *Perspectives on Psychological Science*, 1–14.

Hsee, C. K. (1996), 'The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives', *Organizational Behavior and Human Decision Processes*, **67**(3): 247–257.

Hsee, C. K. (1998), 'Less is better: When low-value options are valued more highly than high-value options', *Journal of Behavioral Decision Making*, **11**, 107–121.

Hsee, C. K., G. F. Loewenstein, S. Blount and M. H. Bazerman (1999), 'Preference reversals between joint and separate evaluations of options: A review and theoretical analysis', *Psychological Bulletin*, **125**(5): 576–590.

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

Hsee, C. K. and J. Zhang (2004), 'Distinction bias: Misprediction and mischoice due to joint evaluation', *Journal of Personality and Social Psychology*, **86**(5): 680–695.
Irwin, J. R., P. Slovic, S. Lichtenstein and G. H. McClelland (1993), 'Preference reversals and the measurement of environmental values', *Journal of Risk and Uncertainty*, **6**(1): 5–18.
Jachimowicz, J. M., S. Duncan, E. U. Weber and E. J. Johnson (2018), When and Why Defaults Influence Decisions: A Meta-Analysis of Default Effects. *Unpublished manuscript*.
Johnson, E. J. and D. Goldstein (2003), 'Do defaults save lives?', *Science*, **302**(5649): 1338–1339.
Jung, J. Y. and B. Mellers (2016), 'American attitudes toward nudges', *Judgment and Decision Making*, **11**(1): 62–74.
Kahneman, D. (2011), *Thinking, fast and slow*, New York, NY: Farrar, Straus and Giroux.
Kahneman, D. and D. T. Miller (1986), 'Norm theory: Comparing reality to its alternatives', *Psychological Review*, **93**(2): 136–153.
Kahneman, D. and I. Ritov (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.
Kahneman, D., I. Ritov and D. A. Schkade (1999), 'Economic preferences or attitude expressions? An analysis of dollar responses to public issues', *Journal of Risk Uncertainty*, **19**, 203–235.
Loewenstein, G., C. Bryce, D. Hagmann and S. Rajpal (2015), 'Warning: You are about to be nudged', *Behavioral Science & Policy*, **1**(1): 35–42.
Milkman, K. L., J. Beshears, J. J. Choi, D. Laibson and B. C. Madrian (2011), 'Using implementation intentions prompts to enhance influenza vaccination rates', *Proceedings of the National Academy of Sciences*, **108**(26): 10415–10420.
Nickerson, D. W. and T. Rogers (2010), 'Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making', *Psychological Science*, **21**(2): 194–199.
Office of the Press Secretary, (2015), Executive order – Using behavioral science insights to better serve the American people. Retrieved on December 28th, 2016 from: https://www.whitehouse.gov/the-press-office/2015/09/15/executive-order-using-behavioral-science-insights-better-serve-american
Reisch, L. A. and C. R. Sunstein (2016), 'Do Europeans like nudges?', *Judgment and Decision Making* **11**(4): 310–325.
Rozin, P., S. Scott, M. Dingley, J. K. Urbanek, H. Jiang and M. Kaltenbach (2011), 'Nudge to nobesity I: Minor changes in accessibility decrease food intake', *Judgment and Decision Making*, **6** (4): 323–332.
Shafir, E. (1999), 'Philosophical intuitions and cognitive mechanisms', in M. R. DePaul and W.M. Ramsey (eds.), *Rethinking intuition: The psychology of intuition and its role in philosophical inquiry*, Lanham, MD, US: Rowman & Littlefield, 59–73.
Shaw, T. (2017), Invisible manipulators of your mind. *The New York Review of Books*. Retrived from: http://www.nybooks.com/articles/2017/04/20/kahneman-tversky-invisible-mind-manipulators/
Steffel, M., E. F. Williams and R. Pogacar (2016), 'Ethically Deployed Defaults: Transparency and Consumer Protection Through Disclosure and Preference Articulation', *Journal of Marketing Research*, **53**(5): 865–880.
Sunstein, C. R. (2016a), 'Do People Like Nudges?', *Administrative Law Review*.
Sunstein, C. R. (2016b), 'People prefer System 2 nudges (kind of)', *Duke Law Journal*, **66**(1): 121–168.
Tannenbaum, D., C. R. Fox and T. Rogers (2017), 'On the misplaced politics of behavioural policy interventions', *Nature Human Behaviour*, **1**(7).
Thaler, R. H. and S. Benartzi (2004), 'Save more tomorrow™: Using behavioral economics to increase employee saving', *Journal of Political Economy*, **112**(S1): S164-S187.
Thaler, R. H. and C. R. Sunstein (2008), *Nudge: Improving Decisions About Health, Wealth, and Happiness*, New Haven, CT: Yale University Press.

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press

Thorndike, A. N., J. Riis, L. M. Sonnenberg and D. E. Levy (2014), 'Traffic-light labels and choice architecture: Promoting healthy food choices', *American Journal of Preventive Medicine*, **46** (2): 143–149.

Van Zant, A. B. and D. A. Moore (2015), 'Leaders' use of moral justifications increases policy support', *Psychological Science*, **26**(6): 934–943.

Waldron, J. (2014), It's all for your own good. *The New York Review of Books*. Retrieved from: http://www.nybooks.com/articles/2014/10/09/cass-sunstein-its-all-your-own-good/

Walton, G. M. and G. L. Cohen (2011), 'A brief social-belonging intervention improves academic and health outcomes of minority students', *Science*, **331**(6023): 1447–1451.

Wansink, B. (2006), *Mindless eating: Why we eat more than we think*, New York, NY: Bantam Dell.

https://doi.org/10.1017/bpp.2018.9 Published online by Cambridge University Press