Behavioural Public Policy, Page 1 of 21
 doi:10.1017/bpp.2018.1

The role of perceived effectiveness on the acceptability of choice architecture

H. MIN BANG*

Fuqua School of Business, Duke University, Durham, NC, USA

SUZANNE B. SHU

Anderson School of Management, University of California at Los Angeles, Los Angeles, CA, USA

ELKE U. WEBER

Department of Psychology, Princeton University, Princeton, NJ, USA

Abstract: The success of choice architecture, including its adoption in government policy and practice, has prompted questions of whether choice architecture design decisions are sufficiently transparent and publicly acceptable. We examined whether disclosing to decision-makers that a particular choice architecture is in place reduces its effectiveness and whether an understanding of the effectiveness of choice architecture design decisions increases their acceptability. We find that disclosure of the design decision does not reduce its effectiveness and that individuals perceive the effectiveness of specific designs to be higher for others than for themselves. Perceived effectiveness for self increases when individuals have actually experienced the effect of a design decision rather than having it simply described to them. Perceived effectiveness for oneself and others increases the acceptability of the designs. We also find that the intentions of the source matter more than who the source actually is. Important for policy-makers, then, is that disclosure of design decisions does not reduce their effectiveness, and their acceptability depends on their perceived effectiveness and the inferred motivations of the design architect.

Submitted 3 May 2017; accepted 9 January 2018

Introduction

From the global success of books like *Thinking Fast and Slow* (Kahneman, 2013) and *Nudge* (Thaler & Sunstein, 2008) to the establishment of "nudge units" in governments of the UK, the USA and many other countries

* Correspondence to: H. Min Bang, Fuqua School of Business, Duke University, Durham, NC 27708, USA. Email: hb71@duke.edu

(OECD, 2014), there is no question about it: choice architecture has arrived and is here to stay! Choice architecture enables policy-makers to structure decision-making environments in ways that promote choices that are in both the decision-makers' and society's interests. Tools include but are not restricted to setting default options or reframing choice options. They also include means that allow decision-makers to focus scarce attention and processing resources on important information and to help with information integration (Johnson *et al.*, 2012). It should be noted that these tools are not necessarily "interventions," since an intervention implies that there is a neutral approach from which the intervention deviates. The literature on choice architecture makes clear that there is no neutral design (Thaler & Sunstein, 2003). Regardless of whether or not the policy-maker intends to encourage a certain behavior over others, meta-decisions about how a choice is presented still have an effect. Given the broad set of tools available and their distinction from traditional interventions, and setting aside the policy-maker's desire to "nudge" some behaviors over others, we use the term "design decisions" for the tools explored within this paper.[1]

With prominence and success comes greater scrutiny. Nudging has been equated to the "manipulation of choice," with calls to make choice architecture design decisions more transparent (Hansen & Jespersen, 2013). The ethics of nudging have been examined both from the inside (Thaler & Sunstein, 2003) and the outside (Raihani, 2013; Bruns *et al.*, 2016). Some researchers have started to examine public attitudes toward nudges (Felsen *et al.*, 2013; Sunstein, 2015; Jung & Mellers, 2016; Tannenbaum *et al.*, 2016), finding greater support for overt nudges (posting nutritional content in a cafeteria) than covert nudges (moving the unhealthy food to an inconvenient location), but leaving many questions (including the domain-specificity of results and the identity of the choice architect) unanswered.

People's awareness of the effectiveness of choice architecture design decisions and the consequences of perceived effectiveness are also poorly understood and under-investigated. In the context of health insurance decisions related to the Affordable Care Act, consumers have shown poor intuitions about which decision environment will help them to optimize their choices (Johnson *et al.*, 2013). This means that they need help in "deciding how to decide" when given the choice between different choice architecture environments. If informed selection of decision aids or informed consent for choice architecture designs is the goal of policy-makers, then we need to better

1 We thank Eric J. Johnson for coining this term and recommending its use.

understand how the public reacts to choice architecture design decisions when they receive information about their presence and likely effect.

The set of studies in this paper was designed to fill this gap and answer the following questions about the consequences of providing decision-makers with information about the presence and influence of a particular design decision: does such knowledge reduce the designs' effectiveness, because decision-makers react against such influence or does it leave the effect intact, as is the case for visual illusions? Do people have different expectations about how the designs may influence their own decisions vs. the decisions of others? Does it matter *how* the effectiveness of a design decision is being disclosed to people (e.g., from personal experience by being exposed to two different versions of the same question or by a simple description of the effect)? Lastly, does the identity of the decision architect and his or her motivation influence the design's acceptability? To examine the generalizability of answers to these questions, we exposed decision-makers to several commonly used types of choice architecture design decisions, including attribute and outcome framing and the use of choice defaults. Specifically, we focus on non-transparent design decisions that influence preferences.

Does choice architecture transparency reduce its effectiveness?

A sense of epistemic transparency has been suggested as an important distinction between design decisions and manipulation. That is, when the intention behind a design decision and the process by which it affects preferences are not easily apparent to decision-makers, the design decision may be considered to be a case of manipulation (Hansen & Jespersen, 2013). Thus, increasing transparency involves a sense of agency that enables decision-makers to recognize and uncover how and why a design decision influences preferences (Bovens, 2009). Disclosure of the effects of choice architecture designs and the subsequent collection of some form of informed consent when exposing individuals to a particular form of choice architecture are obvious solutions to the charge of covert manipulation. The question is whether disclosure may reduce its effectiveness in shaping decision-makers' choices.

Transparency has been defined in different ways in the choice architecture literature. Thaler and Sunstein (2008) equate it with Rawls' publicity principle (Rawls, 1971), which bans governments from implementing any policy they would not be willing and able to defend publicly to their citizens. Later definitions have focused on whether or not the individual being nudged is consciously aware of the presence and details of the design decision, finding mixed results about whether overt designs (of which decision-makers are consciously aware) are preferred over covert (below consciousness) designs

(LeBoeuf & Shafir, 2003; Bruns *et al.*, 2016; Jung & Mellers, 2016). Regardless of the definition of transparency or disclosure, the concern remains that many choice architecture tools may cease to have an effect once the individual is made aware of them.

On the other hand, even after the presence and process of design decisions are made transparent by either describing or experiencing their effects, the effect of choice architecture tools may well remain intact, similar to what happens with visual illusions: just as people still experience visual illusions after they know how they work, individuals who are made aware of the presence and process of a design decision may still be susceptible to its influence. Thus, we predict that the effect of a choice architecture design decision will not be in conflict with its transparency.

What affects a choice architecture design decision's acceptability?

Perceived effectiveness

Although individuals' acceptance of questionable public policies or marketing practices generally increases with their perceived effectiveness, the degree to which perceived effectiveness leads to acceptability often depends on contextual factors and individual differences (e.g., Charry *et al.*, 2014). Especially in situations in which individuals' short-term interests are at odds with long-term benefits, perceived effectiveness may not always influence acceptability; people can understand that long-term benefits leave them better off, but still want to override such a long-term preference with a short-term alternative. In this research, we propose two important factors that affect the relationship between perceived effectiveness and acceptability – a self vs. other discrepancy and the source of the design decision.

Self vs. other discrepancy

Behavioral decision research has documented a wide range of situations in which individuals' judgments for themselves differ from their judgments for others, whether it be believing that others are less risk averse (Hsee & Weber, 1997) or are more average at driving and other tasks (Dunning *et al.*, 1989) or believing that others' actions reflect personality characteristics more than situational influences (Jones & Harris, 1967). More recent research has documented that individuals perceive others to be more vulnerable than they are to judgmental biases, a personal "bias blind spot" (Pronin *et al.*, 2002; Pronin *et al.*, 2004). This gap in understanding one's own susceptibility to bias relative to others' suggests that individuals may also believe that choice architecture will affect the choices of others more than their own.

This self–other discrepancy has important implications for public reactions toward choice architecture design decisions: people tend to deny the presence of social influences on themselves while they overestimate their impacts on others (Robinson *et al.*, 1995; Davison, 1983; Cohen, 2003; Pronin, 2009). For example, people tend to view themselves as less persuaded by the media than other people (Davison, 1983). Such a sense of illusion, however, can be reduced by mere exposure, by which individuals directly experience internal reactions (Gunther & Thorson 1992). Thus, our prediction is grounded in this notion that people acknowledge influences on themselves through experiences. We predict that this divergence in perceptions of oneself and others can be reduced by direct exposure to choice architecture design decisions.

Source and intention behind the design decision

Reaction to any choice architecture design decision will depend partly on how much the decision-maker trusts the intentions and motivations of the choice architect (Bruns *et al.*, 2016; Jung & Mellers, 2016; Tannenbaum *et al.*, 2016). Design architects (i.e., the source of design decisions) can differ in their reputation and motivation for implementing the design, which can affect the perceived fairness of their actions (Campbell, 1999). For example, people tend to perceive defaults as implicit recommendations of the preferred option by policy-makers (McKenzie *et al.*, 2006). Decision-makers who are aware of a source's self-serving motivation to persuade them to choose in a specific way will be especially wary of any design decision by that source (Campbell & Kirmani, 2000).

Building upon the importance of trust and intentions, we expect that type of intentions influences the perceived effectiveness and acceptability of that design decision. The role of perceived intention is also pertinent to the issue of increasing transparency. Given that potential criticism of a choice architecture design mainly stems from the fact that it is intentional (e.g., Bovens, 2009; Wilkins, 2013), understanding how people actually infer the intentions behind the implementation will be linked to acceptability. Specifically, designs decisions for collective wellbeing, such as sustainability or public health reasons, will increase acceptability. To test this hypothesis, we use a set of design decisions that has been suggested as non-transparent (Hansen & Jespersen, 2013).

Research overview

In this paper, we increase transparency by presenting participants with different versions of the same task and asking for people's predictions of the effectiveness of these choice designs for their own decisions or those of others after disclosure

of their presence and influence. We also investigate the effect of the design's perceived effectiveness and the intentions of the source on its acceptability. To do so, we present participants with descriptions of several design decisions and assess perceived effectiveness for both self and others for the designs implemented by different sources (i.e., a government, a company or a friend) and examined whether perceived effectiveness for self or for others positively predicts how acceptable the designs were seen to be. Furthermore, we varied our disclosure methods (description vs. experience) to understand their effects on its effectiveness. Finally, we directly explored intentions by asking respondents what they thought the source's intentions were for implementing the design, which we then analyzed to see its effect on perceived effectiveness and acceptability.

Study 1: perceived effectiveness for self and others

Methods

This study was run on Amazon's Mechanical Turk with 249 participants (53% female, 47% male). Two hundred and forty-two participants passed an attention filter at the start of the survey and were retained for further analysis. The mean age was 36.5 years (SD = 12.2, range = 18–69 years), median income was $40,000–$49,999, 67% were currently employed, 45% had a college degree and self-reported political party affiliation was 40% Democrat, 17% Republican, 32% independent and 11% other or no affiliation.

The choice architecture literature includes a long list of possible design decisions that could be tested for effectiveness and acceptability (see Johnson *et al.*, 2012). For simplicity, we examined three common choice architecture design decisions taken from the existing literature: a design to reduce overeating by using smaller plates (Wansink, 2010); a design to describe the carbon surcharge as a carbon offset (Hardisty *et al.*, 2010); and a change in default settings designed to increase organ donation rates (Johnson & Goldstein, 2003; see Online Appendix for details). Although these design decisions have been suggested as mostly non-transparent nudges (Hansen & Jespersen, 2013), our interest is not in the specific designs, but in how effectiveness and acceptability vary according to description and source. For each of the designs, we disclosed its likely effect by a verbal description. We also varied the described source of the design decision. For the smaller plate design, we specified the source as either a local buffet restaurant, a government cafeteria or a close friend. For the carbon fee framing, we specified the source as either the individual airline selling the ticket or the federal government. For the organ donation default, we described the source as either a health insurance company or a government-sponsored organ donation network.

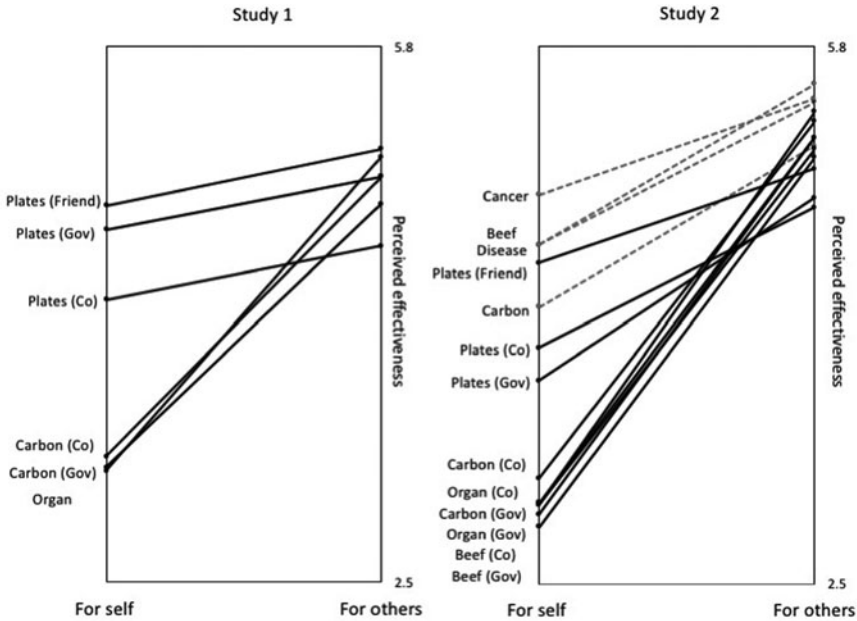

Figure 1. Perceived effectiveness for others is consistently higher than perceived effectiveness for self. Dashed lines represent experienced framings. "Gov" indicates a government source and "Co" indicates a corporate source.

After reading each decision scenario, with its described effect of changes in frame on decisions, participants indicated their level of agreement on a seven-point scale for two statement: "the change will affect my behavior" and "the change will affect others." The first question is designed to measure how effective the individual expects the design to be for him or herself, while the second measures perceived effectiveness for others. We then assessed acceptability of the described change in framing as agreement with the statement, "I am glad that [source] made this change," expressed on a seven-point scale (7 = strongly agree, 1 = strongly disagree).

Results

As shown in Figure 1, participants expected the change to have a much larger effect on others than on themselves. This asymmetry in ratings for self vs. others was statistically significant for each individual task (smaller plates – buffet: $M_{self} = 4.24$, $SD = 1.82$ vs. $M_{others} = 4.57$, $SD = 1.64$, $t(241) = 4.27$, $p < 0.001$; smaller plates – government: $M_{self} = 4.67$, $SD = 1.62$ vs. $M_{others} = 5.00$, $SD = 1.39$, $t(241) = 4.78$, $p < 0.001$; smaller plates – friends: $M_{others} =$

Table 1. A regression of the perceived effectiveness for self and others, controlling for tasks and sources.

	Model 1		Model 2	
	b	*SE* b	b	*SE* b
Study 1 (described framings only)				
Other	1.04***	0.06	1.04***	0.06
Plates task	0.59***	0.08	0.18***	0.10
Carbon task	−0.07	0.08	0.30**	0.08
Corporate source			−0.45***	0.09
Government source			−0.30***	0.09
Adjusted R^2	0.42		0.42	
Study 2 (both experienced and described framings)				
Other	1.51***	0.04	1.51***	0.04
Plates task	0.11	0.07	0.19*	0.08
Carbon task	−0.04	0.07	−0.08	0.07
Beef task	0.01	0.07	−0.03	0.07
Disease task	0.21	0.11	0.08	0.11
Cancer task	0.37***	0.11	0.25*	0.11
Corporate source	−0.63***	0.06	−0.38***	0.10
Government source	−0.70***	0.06	−0.45***	0.10
Experience			0.37**	0.11
Adjusted R^2	0.36		0.36	

"Other" is included as a binary indicator for whether effectiveness is judged for the self or for others. "Experience" is included as a binary indicator for whether participants experienced the tasks in Study 2. Friend source and organ donation task were omitted because of collinearity. Coefficients are unstandardized.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

5.17, SD = 1.48; M_{self} = 4.82, SD = 1.78, $t(241) = 5.00$, $p < 0.001$; organ donation defaults: M_{self} = 3.19, SD = 2.15 vs. M_{others} = 5.12, SD = 1.55, $t(241) = 13.86$, $p < 0.001$; carbon fee – airline: M_{self} = 3.28, SD = 2.05 vs. M_{others} = 5.00, SD = 1.46, $t(241) = 13.56$, $p < 0.001$; carbon fee – government: M_{self} =3.21, SD = 2.05 vs. M_{others} = 4.83, SD = 1.56, $t(241) = 12.95$, $p < 0.001$). In a regression of perceived effectiveness as a function of target (self or other), controlling for both design type (smaller plates, carbon fee and organ donation) and source (company or government) and with clustered errors at the individual level to account for repeated measures per participant, perceived effectiveness for others was significantly larger than for the self (b = 1.04, SE = 0.06, $t = 18.97$, $p < 0.001$; see Table 1).

We next examined whether the perceived effectiveness of each design affected its acceptability. A regression of the acceptability measure against

the perceived effectiveness measures, again controlling for design and source, revealed that higher perceived effectiveness for others exerted a positive influence on how acceptable a design decision was seen to be ($b_{other} = 0.27$, SE = 0.04, $t = 6.55$, $p < 0.001$), whereas higher perceived effectiveness for oneself does not predict acceptability ($b_{self} = 0.04$, SE = 0.03, $t = 1.16$, $p = 0.24$). Thus, while increases in effectiveness can increase acceptability, not all increases in effectiveness are equally impactful.

Study 2: experience vs. description

The goal of Study 2 was to replicate the previous result, but also to broaden our list of examined designs as well as our disclosure method. As with Study 1, our interest was not in the specific designs, but in how changes in description and source affect perceived effectiveness and acceptability. For half of the decision scenarios, we verbally described the effect of the change in frame/design as in Study 1. For the other half of the decision scenarios, we disclosed the effect of the change in frame by exposing participants to both versions, allowing them to experience the effect on their own choices and judgments. We expected individuals who experience two different frames to be more aware of the effectiveness of the frame/design on their own choices. In addition, we more deeply explored the source effects by looking at how differences in perceived intentions affected the acceptability of each design decision per source.

Methods

Our data come from an online sample of 226 Amazon Mechanical Turk participants (54% female, 46% male) who had not participated in the previous study. Our total survey population was 242 individuals, but only the 93.4% who passed the attention filter are included in the analysis. All participants were US residents. Median income was $30,000–$39,999, 35% were currently married, 65% were currently employed, 38% had a college degree and self-reported political party affiliation was 42% Democrat, 17% Republican, 24% independent and 17% other or no affiliation.

The four message framing tasks for which the effect of frame was personally experienced were the Asian Disease problem (Tversky & Kahneman, 1981), lean beef (Braun *et al.*, 1997), carbon fees (Hardisty *et al.*, 2010) and cancer treatment options (McNeil *et al.*, 1982). All four used valence framing, in which outcomes are described in positive terms or negative terms, but two focused on risky choice and changed the numeric description of choice options (Asian Disease and cancer treatment), while the other two focused on riskless choice and changed the description of outcome attributes (beef

and carbon fee). The Asian Disease task described two treatment options for a disease outbreak with outcomes framed as either lives saved (positive) or lives lost (negative) and measured choice for the risky vs. the certain option. The lean beef task described grocery store ground beef as either 75% lean (positive) or 25% fat (negative) and measured judgments of quality (taste, greaseless, quality and lean)[2] and purchase likelihood. The carbon fee task described an additional airline fee as being either a carbon offset (positive) or a carbon tax (negative) and measured choice for an airline ticket with or without the fee. Finally, the cancer treatment task described lung cancer treatment outcomes in terms of survival (positive) or mortality (negative) and measured choice between surgery and radiation therapy.

Participants experienced the effects of framing design by first making choices or judgments for options described using one frame and then again for the same options described using the alternative frame. The connection between the two frames was not disguised, but the two versions of the decision were separated in time and their connection was not made explicit until after all questions were completed. The order of the two frames for each task was counterbalanced across subjects. After participants indicated their answers for both versions of all four tasks, they were each shown the scenario again and were informed that the two frames were in fact equivalent. At that point, individuals who had given inconsistent answers between the two frames were asked if they would like to change one of their answers. After giving them the chance to change any inconsistent answers, we measured perceived effectiveness for both self and others. Participants indicated their level of agreement on a seven-point scale for the same statements: "the change will affect my behavior" and "the change will affect others."

The design decisions in the next four tasks were described rather than experienced. Two were variations of scenarios participants had experienced previously: beef framing and carbon fee framing. Two tasks were new: a design to reduce overeating by using smaller plates and a change in default settings designed to increase organ donation rates. As in Study 1, we varied the described source of the design decision, using the same sources as in Study 1. For the beef framing design, we specified the source of the design decision as either the local grocery store or the federal government. For the organ donation default settings, we described the source of the change as either a medical insurance company or the federal government. We assessed perceived effectiveness for both self and others using the same two questions ("the change will affect my behavior" and "the change will affect others") and then assessed

2 These four judgment variables are combined into a single overall quality measure ($\alpha = 0.84$).

the acceptability of the changes in frame/designs: "I am glad that [source] made this change."

To measure perceived motivation for each source, we provided participants with a drop-down list of possible motivations and also allowed free text entry. The motivations were then coded into categories of financial or profit-related reasons (e.g., "keep costs lower and increase profits," "increase sales/revenue," "avoid taxes"), sustainability reasons ("reduce waste," "prevent global warming," "reduce overconsumption"), health-related reasons ("promote healthy eating," "save lives") or other reasons ("to set a good example," "for control").

Results

When we consider each participant's response to the first frame as a between-subject test of outcome framing, we replicate the standard framing effects for each scenario. In the Asian Disease problem, 22% of respondents choose the risky option in the gain frame vs. 71% in the loss frame ($\chi^2 = 53.7$, $p < 0.001$); this compares to 28% and 78%, respectively, in the original study. Similar results are found for each of the other three tasks. For the beef framing task, individuals who first saw the beef described as 25% fat perceive it as lower quality and are (directionally but not significantly) less likely to purchase it than those who first saw it described as 75% lean (overall quality: 3.5 vs. 4.31, $t(224) = 4.84$, $p < 0.001$; purchase: 3.0 vs. 3.32, $t(224) = 1.48$, $p = 0.14$). For the carbon fee framing task, 54% of individuals chose to pay the extra fee when they first saw it described as an offset, but only 38% do when they first saw it described as a tax ($\chi^2 = 5.10$, $p = 0.02$). This compares to 60% and 39%, respectively, in the original study. For the cancer treatment task, individuals in both frames have a strong preference for radiation over surgery, but that preference shrinks from a 71% preference when surgery results are framed in terms of death to a 56% preference when surgery results are framed in terms of survival ($\chi^2 = 4.73$, $p = 0.03$), consistent with the pattern of framing results originally found by McNeil *et al.* (1982). The effects of frame on choices or judgments for all tasks are summarized in Supplementary Online Materials.

In the Asian Disease problem, 57% of participants who first saw the gain frame and 65% of participants who first saw the loss frame stayed consistent with their initial choice; of the inconsistent individuals, 89% (39%) were inconsistent in the direction predicted by prospect theory (Kahneman & Tversky, 1979). These proportions are similar to previous findings on consistency in within-subject versions of this task (Stanovich & West, 1998; LeBoeuf & Shafir, 2003). Among the 88 inconsistent individuals, only 46 indicated that

they would like to change one of their answers. For the other three tasks, we find even stronger results. Participants choosing airline tickets with carbon offsets stayed consistent with their initial choice across frames 96% of the time and participants choosing cancer treatments stayed consistent 80.5% of the time. Less than half of inconsistent individuals wanted to change their answers once their inconsistency was made salient.[3]

For each of the four scenarios, perceived effectiveness of the change in frame is judged consistently higher for others than for oneself (Asian Disease $M_{self} = 4.58$, $SD = 2.01$ vs. $M_{other} = 5.46$, $SD = 1.42$, $t(225) = 8.78$, $p < 0.001$; beef framing: $M_{self} = 4.58$, $SD = 2.23$ vs. $M_{other} = 5.57$, $SD = 1.41$, $t(225) = 7.59$, $p < 0.001$; carbon fee: $M_{self} = 4.20$, $SD = 2.28$ vs. $M_{other} = 5.19$, $SD = 1.75$, $t(225) = 8.32$, $p < 0.001$; and cancer treatment $M_{self} = 4.89$, $SD = 1.95$ vs. $M_{other} = 5.48$, $SD = 1.38$, $t(225) = 5.89$, $p < 0.001$). However, we also find that perceived effectiveness differs when participants gave consistent or inconsistent answers across the two frames in these tasks. Recall, for example, that 88 participants gave inconsistent answers on the Asian Disease problem. Compared to consistent respondents, the inconsistent respondents' perceived effectiveness is significantly higher for the self and directionally higher for others (5.25 vs. 4.15, $t(224) = 4.13$, $p < 0.001$ for self; 5.61 vs. 5.37, $t(224) = 1.26$, $p = 0.20$ for others). Among the group of inconsistent respondents, perceived effectiveness for self is still significantly lower than for others ($M_{self} = 5.25$ vs. $M_{other} = 5.61$, $t(87) = 3.79$, $p < 0.001$), but the size of this self–other gap is significantly smaller than for the consistent individuals (1.22 for consistent vs. 0.36 for inconsistent, $t(224) = 4.28$, $p < 0.001$). The size of the self–other gap for individuals making inconsistent choices is even smaller for the other two message framing tasks,[4] to the point of being no longer significant (carbon fee: $M_{self} = 5.33$ vs. $M_{other} = 5.67$, $t(8) = 0.48$, non-significant; cancer treatment: $M_{self} = 5.20$ vs. $M_{other} = 5.31$, $t(43) = 0.96$, non-significant).

Consistent with Study 1, the gap between perceived effectiveness for self vs. others is apparent for the design decisions that were described rather than experienced (smaller plates – buffet: $M_{self} = 3.95$, $SD = 2.01$ vs. $M_{others} = 4.82$, $SD = 1.60$, $t(224) = 8.05$, $p < 0.001$; smaller plates – government: $M_{self} = 3.75$, $SD = 1.99$ vs. $M_{others} = 4.88$, $SD = 1.55$, $t(224) = 9.57$, $p < 0.001$; smaller plates – friends: $M_{self} = 4.47$, $SD = 1.87$ vs. $M_{others} = 5.05$, $SD = 1.51$, $t(224) = 5.78$, $p < 0.001$; carbon fee – airline: $M_{self} = 3.00$, $SD = 2.15$ vs. $M_{others} = 5.24$,

3 A more detailed analysis of inconsistent individuals from these two tasks is not meaningful due to the small number of inconsistent subjects in each group (9 for carbon offsets, 44 for cancer treatments).

4 Recall that only three tasks were choice tasks (the beef framing design was a judgment task), so analyses of consistent vs. inconsistent results are only available for these three.

SD = 1.46, $t(24) = 15.57$, $p < 0.001$; carbon fee – government: M_{self} =2.99, SD = 2.08 vs. M_{others} = 5.17, SD = 1.57, $t(224) = 16.23$, $p < 0.001$; organ donation – company: M_{self} = 3.00, SD = 2.15 vs. M_{others} = 5.24, SD = 1.46, $t(224) = 15.41$, $p < 0.001$; organ donation – government: M_{self} = 2.99, SD = 2.19 vs. M_{others} = 5.40, SD = 1.46, $t(224) = 15.73$, $p < 0.001$; beef – store: M_{self} = 2.93, SD = 2.05 vs. M_{others} = 5.24, SD = 1.39, $t(224) = 16.87$, $p < 0.001$; beef – government: M_{self} = 2.85, SD = 2.00 vs. M_{others} = 5.13, SD = 1.46, $t(224) = 17.17$, $p < 0.001$).

As in Study 1, a regression of perceived effectiveness on self vs. other, controlling for both design and source and with clustered errors at the individual level to account for repeated measures per participant, showed that perceived effectiveness for others was significantly larger than for the self (b = 1.51, $t = 35.53$, $p < 0.001$). Since we measured perceived effectiveness for both experienced (Asian Disease, carbon fee, beef labeling and cancer treatment) and described designs (smaller plates, carbon fee, organ donation and beef labeling), we examined whether the type of disclosure affects perceived effectiveness. Results indicate that that perceived effectiveness for experienced designs is significantly higher than for described designs, even when controlling for the type of tasks (b = 0.37, SE = 0.11, $t = 3.26$, $p = 0.001$).

A regression of the gap between perceived effectiveness for self and for others on the disclosure methods revealed that the difference between self and others was significantly smaller for experienced design decisions than for described ones, even when controlling for tasks and sources (b = –0.85, SE = 0.15, $t = -5.55$, $p < 0.001$; see Table 2).

Perceived intentions for each design decision differed by source. For the carbon fee framing, both the airline and the government were seen as mostly being guided by financial motivations (51% and 41%, respectively), suggesting that individuals suspect profit motives; however, the government received a higher number of responses for sustainability issues such as "prevent global warming" (46.7% vs. 35.6% for industry). Differences are even greater for the smaller plates design decision. Here, financial motives (cost and/or profit) were seen as the dominant motivation for industry, at 77% of responses. For the government cafeteria and the friend, however, the majority response was "promote healthier eating," at 47.1% and 48.4%, respectively. For the organ donation and beef framing design decisions, financial motivations were perceived to be higher for companies (80% vs. 23.1% for beef framing), but health motivations were perceived to be higher for governments (74.2% vs. 50.2% for organ donation defaults). Perceived intentions for all of the tasks are summarized in Supplementary Online Materials.

The source of the design decisions differently affects participants' perceived effectiveness for themselves and their judgments of effectiveness for others. As shown in Table 3, when participants rated design decisions' effectiveness for

Table 2. A regression of the self–other discrepancy against the disclosure method (experience vs. description), controlling for tasks and sources in Study 2.

	Model 1		Model 2	
	b	*SE* b	b	*SE* b
Experience	−1.26***	0.09	−0.85***	0.15
Plates task	−1.47***	0.09	−1.33***	0.10
Carbon task	−0.13	0.10	−0.13	0.10
Beef task	−0.05	0.10	−0.05	0.10
Disease task	−0.19	0.15	−0.19	0.15
Cancer task	−0.49**	0.15	−0.49**	0.15
Corporate source			0.37**	0.13
Government source			0.46***	0.13
Adjusted R^2	0.45		0.45	

"Experience" is included as a binary indicator for whether participants experienced the tasks. Friend source and organ donation task were omitted because of collinearity. Coefficients are unstandardized.
**$p < 0.01$, ***$p < 0.001$.

themselves, design decisions that were implemented by a government source or a corporation source were seen as less effective relative to a friend ($b_{corporate}$ = –0.42, SE = 0.15, t = –2.89, p = 0.004 and $b_{government}$ = –0.60, SE = 0.14, t = –4.33, $p < 0.001$). However, the source of designs did not affect perceived effectiveness for others ($b_{corporate}$ = –0.08, SE = 0.11, p = 0.45 and $b_{government}$ = –0.16, SE = 0.10, p = 0.12). For both effectiveness ratings, sustainability intentions increased perceived effectiveness, holding constant design source (for self: $b_{sustainability}$ = 0.40, SE = 0.17, t = 2.44, p = 0.01; for others: $b_{sustainability}$ = 0.24, SE = 0.12, t = 2.02, p = 0.04).

A regression of the acceptability measure against the perceived effectiveness measures, controlling for design and source, indicated that higher perceived effectiveness for oneself and higher perceived effectiveness for others both predicted acceptability of the designs (b_{self} = 0.21, SE = 0.03, t = 7.79, $p < 0.001$; b_{other} = 0.36, SE = 0.04, t = 9.93, $p < 0.001$). As shown in Table 4, there were also significant effects of source in our regression analyses of acceptability, even when controlling for the effects of the two perceived effectiveness measures analyzed above. Across all tasks, design acceptability was lower for both the government and corporations relative to a friend ($b_{corporate}$ = –0.30, SE = 0.14, t = –2.13, p = 0.03 and $b_{government}$ = –0.36, SE = 0.14, t = –2.50, p = 0.01). This relationship especially held for the small plates design decision,

Table 3. A separate regression of the perceived effectiveness for self and others against tasks and sources for described framings in Study 2.

	Model 1		Model 2	
	b	*SE* b	b	*SE* b
Perceived effectiveness for self				
Plates task	0.86***	0.12	0.90***	0.12
Carbon task	−0.08	0.11	0.02	0.13
Beef task	−0.10	0.11	0.00	0.12
Corporate source	−0.56***	0.14	−0.42**	0.15
Government source	−0.68***	0.14	−0.60***	0.14
Financial motive			−0.11	0.13
Sustainability motive			0.40*	0.17
Health motive			0.14	0.13
Adjusted R^2	0.46		0.46	
Perceived effectiveness for others				
Plates task	−0.47***	0.08	−0.42***	0.09
Carbon task	−0.07	0.08	−0.05	0.10
Beef task	−0.13	0.08	−0.04	0.09
Corporate source	−0.20	0.10	−0.08	0.11
Government source	−0.22*	0.10	−0.16	0.10
Financial motive			−0.09	0.10
Sustainability motive			0.24*	0.12
Health motive			0.15	0.10
Adjusted R^2	0.42		0.42	

Friend source, organ donation task and other motives were omitted because of collinearity. Coefficients are unstandardized.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

where acceptability was higher for a friend than for either a buffet restaurant or government cafeteria (4.78 vs. 4.24 and 4.25, $p < 0.001$ for both). Whether a design decision was more acceptable when it originates from the government or a corporation changed with the design. Beef framing was more acceptable when it came from the grocery store than the government (4.09 vs. 3.50, $t(224) = 5.98$, $p < 0.001$), but organ donation defaults were more acceptable when they came from the government than when they came from an insurance company (4.18 vs. 3.8, $t(224) = 3.76$, $p < 0.001$). Carbon fees were acceptable regardless of whether they came from the airline or the government (4.42 vs. 4.30, non-significant).

When comparing acceptability rating by motive across designs and sources, we found that acceptability for design decisions perceived to be done for sustainability ($\bar{x} = 5.09$) and health ($\bar{x} = 4.66$) motivated reasons were significantly

Table 4. A regression of the acceptability measure against the perceived effectiveness measures and perceived intentions for described tasks in Study 2.

	Model 1		Model 2		Model 3	
	b	*SE* b	b	*SE* b	b	*SE* b
Plates task	0.26*	0.12	0.25*	0.12	0.48**	0.13
Carbon task	0.37**	0.12	0.38***	0.11	0.55***	0.14
Beef task	−0.20	0.12	−0.13	0.11	0.20	0.13
Corporate source	−0.49**	0.15	−0.30*	0.14	−0.10	0.16
Government source	−0.57***	0.15	−0.36*	0.14	−0.38	0.15
Effectiveness for self			0.21***	0.03		
Effectiveness for others			0.36***	0.04		
Financial motive					−0.13	0.14
Sustainability motive					0.82***	0.18
Health motive					0.77***	0.15
Adjusted R^2	0.38		0.46		0.40	

Model 1 accounts for the source of the intervention. Model 2 accounts for perceived effectiveness for self and others. Model 3 accounts for perceived intention. Organ donation and friend source were omitted because of collinearity. Coefficients are unstandardized.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

higher than average for those perceived to have either financial ($\bar{x} = 3.72$) or other reasons ($\bar{x} = 3.59$, all $p < 0.001$). Compared to financial motives, both sustainability and health intentions increase acceptability by almost a full point on the seven-point scale ($b_{sustainability} = 0.82$, SE = 0.18, $t = 4.65$, $p < 0.001$ and $b_{health} = 0.77$, SE = 0.15, $t = 5.31$, $p < 0.001$). The full regression with controls for participants' political views and demographics is available in the Supplementary Online Materials.

General discussion

Across multiple choice architecture tools, we explored the effects of two design disclosures (by experience or description) on people's perceptions of their effectiveness and acceptability. We also examined whether design decisions implemented by a government source would be perceived differently than design decisions implemented by a company or by a friend. Our findings have important implications for policy-makers interested in implementing choice architecture design decisions, suggesting how to improve the perceived effectiveness and acceptability of such design decisions.

Perceived effectiveness of the change in frame is judged consistently higher for others than for oneself for all tested tasks in the present research, broken

down by task and source. Regardless of the type of design decision, the bias blind spot leads individuals to expect the design to change other people's behaviors more than their own. Both higher perceived effectiveness for oneself and higher perceived effectiveness for others positively influence how acceptable an intervention is seen to be, holding constant other moderators like design source. Therefore, increasing both perceived self and other effectiveness will lead to greater acceptability, with the additional insight that perceived effectiveness for others can have a larger impact on acceptability when perceived effectiveness for the self is relatively low (Study 1).

Given that people consistently view choice architecture design decisions as less effective for themselves than for others, increasing perceived effectiveness for oneself is important. Our findings provide important insights into this issue. First, direct exposure to design decisions increases perceived effectiveness for oneself. In Study 2, perceived effectiveness is higher when individuals have actually experienced the effect of a design decision, rather than having it simply described. Moreover, compared to consistent respondents, respondents who gave inconsistent answers across the two frames in a task rated perceived effectiveness significantly higher for themselves and directionally higher for others. Individuals who have observed their inconsistent choices and judgments view a design decision as being more effective. Their perceived effectiveness for others is still significantly higher than for self, but the size of this self–other gap is significantly smaller when respondents have a chance to personally experience design decisions. Our findings suggest that people's lack of awareness of their susceptibility to the design decision effects – the bias blind spot – can be attenuated by direct exposure to the effects. Efforts to help individuals experience a design decision before its implementation, even if only hypothetically, can increase support for that design decision. For example, websites that allow consumers to toggle the framing of a fee (offset or tax) or buffets that encourage visitors to try walking through with differently sized plates at different visits may provide individuals with the opportunity to test for themselves how their choices might change under different designs.

Second, perceived effectiveness for oneself is sensitive to the source of design decisions, whereas source has little effect on perceived effectiveness for others. Participants view the design decisions for the personal decisions made in our studies as less effective for themselves when they are implemented by a government or a corporate source rather than by a friend. This pattern was not found for perceived effectiveness for others, relative to the impact on others. Our finding is consistent with the notion that people tend to deny their susceptibility to political ideology or media campaigns, whereas they view others as more susceptible to such influences (Pronin, 2009).

In addition to perceived effectiveness, source of the design decision also significantly affected the acceptability of interventions. Acceptability was significantly lower for both government and corporate sources than for a friend. However, differences in acceptability of designs from government and corporate sources change per design. Beef framing is significantly more acceptable when it is done by the grocery store than by the government, while organ donation defaults are more acceptable when they come from the government than from an insurance company. Carbon fees are acceptable regardless of whether they come from the airline or the government. Thus, there may be no simple rule of thumb about design decision acceptability as a function of source, making it even more important to understand the role of perceived intentions behind the design decision.

Type of intention significantly influenced acceptability, even when controlling for tasks, sources and individual difference measures. Design decisions for financial reasons that benefit the choice architect are less acceptable than those done for sustainability or health reasons. Furthermore, negative effects of corporate source are no longer significant once intentions are included, suggesting that it is not the source per se, but the source's perceived intentions that make a design decision less publicly acceptable. Design decisions for sustainability reasons lead to higher perceived effectiveness for oneself and for others, regardless of their source. People are sensitive to the source of design decisions, but further highlighting sustainability intentions has the potential to improve perceptions of their effectiveness. Thus, making clear the intentions behind a design decision can be a valuable method for increasing overall acceptability when the designer has social good (rather than profit) in mind.

Finally, our findings also suggest that individuals attempt to stay consistent with whichever frame they first experience, as long as they recognize the similarity of the tasks, even after the presence and process of design decisions are made transparent and easily predictable through either the description or experiencing their effects. However, for inconsistent individuals, traditional framing effects persist, presumably because they perceive the two versions be different even after their equivalence is asserted. This finding complements the robustness of recent work on framing effects in hospital environments, which has found choices made under initial message frames highly resistant to change, even after the effect of the frame is made salient (Kahneman & Tversky, 1979; Loewenstein *et al.*, 2014).

Conclusion

The present study offers good news for choice architects. Acceptability was generally high for the broad range of design decisions we tested, significantly

above the midpoint of our scale in most cases. Our findings are consistent with recent surveys that show strong support for many design decisions across the spectrum of political beliefs (Sunstein, 2015). Interestingly, our assessed level of *ex ante* acceptability – that is, acceptability judged for hypothetical design decisions not yet enacted – may underestimate public *ex post* support for such measures, as suggested by recent research that shows that when design decisions are enacted despite public opposition, their acceptability can quickly increase after implementation, as predicted by query theory (Treuer *et al.*, 2012; Weber, 2015). Given that choice architecture and design decisions are unavoidable since there is no value-neutral way of displaying information, policy-makers tasked and entrusted with increasing public welfare would seem to have an obligation to make such policy design decisions carefully, implementing them in ways that encourage efficiency, effectiveness and public acceptance. We offer the results of this program of research as guidance in this endeavor.

Supplementary Material

To view supplementary material for this article, please visit https://doi.org/10.1017/bpp.2018.1.

References

Braun, K. A., G. J. Gaeth, and I. P. Levin (1997), 'Framing effects with differential impact: The role of attribute salience', *Advances in Consumer Research*, **24**: 405–11.

Bruns, H., E. Kantorowicz-Reznichenko, K. Klement, M. Luistro Jonsson, and B. Rahali (2016), 'Can Nudges Be Transparent and Yet Effective?' Available at SSRN: http://ssrn.com/abstract=2816227.

Bovens, L. (2009), 'The Ethics of Nudge', In T. Grüne-Yanoff and S. O. Hansson (eds), *Preference Change: Approaches from Philosophy, Economics and Psychology*, (pp. 207–220). Berlin: Springer.

Campbell, M. C. (1999), '"Why did you do that?" The important role of inferred motive in perceptions of price fairness', *Journal of Product and Brand Management*, **8**(2): 145–53.

Campbell, M. C., and A. Kirmani (2000), 'Consumers' use of persuasion knowledge: The effects of accessibility and cognitive capacity on perceptions of an influence agent', *Journal of Consumer Research*, **27**(1): 69–83.

Charry, K., P. Pelsmacker, and C. L. Pecheux (2014), 'How does perceived effectiveness affect adults' ethical acceptance of anti-obesity threat appeals to children? When the going gets tough, the audience gets going', *Journal of Business Ethics*, **124**(2): 243–257.

Cohen, G. L. (2003), 'Party over policy: The dominating impact of group influence on political beliefs', *Journal of Personality and Social Psychology*, **85**: 808–822.

Davison, W. P. (1983), 'The third-person effect in communication', *Public Opinion Quarterly*, **47**: 1–15.

Dunning, D., J. A. Meyerowitz, and A. D. Holzberg (1989), 'Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability', *Journal of Personality and Social Psychology*, **57**(6): 1082–90.

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): 202–13.

Gunther, A. C., and E. Thorson (1992), 'Perceived persuasive effects of product commercials and public service announcements: Third-person effects in new domains', *Communication Research*, **19**: 574–596.

Hansen, P. G., and A. M. Jespersen (2013), 'Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behavior change in public policy', *European Journal of Risk Regulation*, **1**: 3–28.

Hardisty, D. J., E. J. Johnson, and E. U. Weber (2010), 'A dirty word or a dirty world?: Attribute framing, political affiliation, and query theory', *Psychological Science*, **21**(1): 86–92.

Hsee, C. K., and E. U. Weber (1997), 'A fundamental prediction error: Self-others discrepancies in risk preference', *Journal of Experimental Psychology: General*, **126**(1): 45–53.

Johnson, E. J., R. Hassin, T. Baker, A. T. Bajger, and G. Treuer (2013), 'Can consumers make affordable care affordable? The value of choice architecture', *PloS one*, **8**(12): e81521.

Johnson, E. J., S. B. Shu, B. G. C. Dellaert, C. Fox, D. G. Goldstein, G. Haeubl, R. P. Larrick, J. W. Payne, D. Schkade, B. Wansink, and E. U. Weber (2012), 'Beyond nudges: Tools of a choice architecture', *Marketing Letters*, **23**: 487–504.

Jones, E. E., and V. A. Harris (1967), 'The attribution of attitudes', *Journal of Experimental Social Psychology*, **3**(1): 1–24.

Jung, J. Y., and B. A. Mellers (2016), 'American attitudes toward nudges', *Judgment and Decision Making*, **11**(1): 62–74.

Kahneman, D. (2013), *Thinking fast and slow*, New York, Farrar: Straus and Giroux.

Kahneman, D., and A. Tversky (1979), 'Prospect Theory: An analysis of decisions under risk', *Econometrica*, **47**: 263–91.

LeBoeuf, R. A., and E. Shafir (2003), 'Deep thoughts and shallow frames: on the susceptibility to framing effects', *Journal of Behavioral Decision Making*, **16**(2): 77–92.

Loewenstein, G., C. Bryce, D. Hagmann, and S. Rajpal (2014), 'Warning: You are about to be nudged'. Available at SSRN: http://ssrn.com/abstract=2417383.

McKenzie, C. R. M., M. J. Liersch, and S. R. Finkelstein (2006), 'Recommendation implicit in policy defaults', *Psychological Science*, **17**(5): 414–420.

McNeil, B., S. Pauker, H. Tversky , Jr., and A. Sox (1982), 'On the elicitation of preferences for alternative therapies', *New England Journal of Medicine*, **306**: 1259–62.

Organisation for Economic Co-operation and Development (OECD). (2014), *Behavioural Insights and New Approaches to Policy Design*, viewed 23 January 2014. http://www.oecd.org/naec/NAEC_Behavioural-Insights-Programme_23-Jan.pdf.

Pronin, E. (2009), 'The Introspection Illusion', In M. P. Zanna (ed), *Advances in Experimental Social Psychology*, (Vol. 41, pp.1–67). Burlington, MA: Academic Press.

Pronin, E., T. Gilovich, and L. Ross (2004), 'Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others', *Psychological Review*, **111**(3): 781–99.

Pronin, E., D. Y. Lin and L. Ross (2002), 'The bias blind spot: Perceptions of bias in self versus others', *Personality and Social Psychology Bulletin*, **28**(3): 369–81.

Raihani, N. J. (2013), 'Nudge politics: Efficacy and ethics', *Frontiers in Psychology*, **4**: 972–5.

Robinson, R. J., D. Keltner, A. Ward, and L. Ross (1995), 'Actual versus assumed differences in construal: "Naïve realism" in intergroup perception and conflict', *Journal of Personality and Social Psychology*, **68**: 404–417.

Stanovich, K. E., and R. F. West (1998), 'Individual differences in rational thought', *Journal of Experimental Psychology: General*, **127**(2): 161–88.

Rawls, J. (1971), *A theory of justice*, Cambridge, Mass: Harvard University Press.
Sunstein, C. R. (2015), 'Do people like nudges?' Available at SSRN: http://ssrn.com/abstract=2604084.
Tannenbaum, D., C. R. Fox, and T. Rogers (2016), 'On the misplaced politics of behavioral policy interventions'. Working paper.
Thaler, R. H., and C. R. Sunstein (2008), *Nudge: Improving decisions about health, wealth, and happiness*, New Haven, CT: Yale University Press.
Thaler, R. H., and C. R. Sunstein (2003), 'Libertarian paternalism', *American Economic Review*, **93**: 175–9.
Treuer, G. A., K. C. Appelt, A. E. Goll, R. D. Crookes, and E. U. Weber (2012), 'Weathering the storm: Status quo adjustments explain successful policy implementation'. Working paper.
Tversky, A., and D. Kahneman (1981), 'The framing of decisions and the psychology of choice', *Science*, **211**: 453–8.
Weber, E. U. (2015), 'Climate change demands behavioral change: What are the challenges?', *Social Research: An International Quarterly*, **82**: 561–81.
Wilkins, T. M. (2013), 'Nudging and manipulation', *Political Studies*, **61**: 341–55.