



## Management Science

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### When Do People Prefer Carrots to Sticks? A Robust “Matching Effect” in Policy Evaluation

Ellen R. K. Evers, Yoel Inbar, Irene Blanken, Linda D. Oosterwijk

To cite this article:

Ellen R. K. Evers, Yoel Inbar, Irene Blanken, Linda D. Oosterwijk (2016) When Do People Prefer Carrots to Sticks? A Robust “Matching Effect” in Policy Evaluation. Management Science

Published online in Articles in Advance 17 Oct 2016

<http://dx.doi.org/10.1287/mnsc.2016.2539>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article’s accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2016, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# When Do People Prefer Carrots to Sticks? A Robust “Matching Effect” in Policy Evaluation

Ellen R. K. Evers

University of California, Berkeley, Berkeley, California 94720, [evers@haas.berkeley.edu](mailto:evers@haas.berkeley.edu)

Yoel Inbar

University of Toronto, Toronto, Ontario M1C 1A4, Canada, [yoel.inbar@utoronto.ca](mailto:yoel.inbar@utoronto.ca)

Irene Blanken, Linda D. Oosterwijk

Tilburg University, 5037 AB Tilburg, Netherlands  
{[ireneblanken@gmail.com](mailto:ireneblanken@gmail.com), [l.d.oosterwijk@tilburguniversity.edu](mailto:l.d.oosterwijk@tilburguniversity.edu)}

For a policy to succeed, it must not only be effective in changing behavior but must also be accepted by stakeholders. Here, we report seven sets of studies demonstrating strong framing effects on the acceptance of equivalent policies. Policies targeting desirable voluntary behavior are preferred when they are framed as *advantaging* those who act desirably (rather than *disadvantaging* those who do not). Conversely, policies targeting obligations are preferred when they are framed as *disadvantaging* those who fail to act desirably (rather than *advantaging* those who do). These differences in policy acceptance do not result from common causes of framing effects, such as a misunderstanding of outcomes or insufficient deliberation about the implications. Rather, the framing effects we document follow from beliefs about when punishment is and is not appropriate. We conclude with a field experiment demonstrating framing effects in a setting where policy acceptance directly affects respondents' outcomes.

Data, as supplemental material, are available at <https://doi.org/10.1287/mnsc.2016.2539>.

**Keywords:** policies; framing; punishing and rewarding; justice; moral decision making

**History:** Received September 29, 2015; accepted March 18, 2016, by Yuval Rottenstreich, judgment and decision making. Published online in *Articles in Advance* October 17, 2016.

*There is nothing whatever objectionable in the idea of someone's being pressed to carry out such a basic duty as promise-keeping, [but] there is something horrifying in the thought of pressure being brought on him to perform an act of heroism.*  
—J. O. Urmson (1958, p. 214)

Policies (rules and regulations intended to affect behavior) can penalize undesirable behavior, reward desirable behavior, or both. For a policy to succeed, it must be both *effective* in changing behavior and *accepted* by stakeholders. A great deal of research has addressed policy effectiveness, testing, for example, the relative efficacy of penalties or rewards (e.g., Andreoni et al. 2003), the effects of social norms (e.g., Cialdini et al. 2006), and status quo biases (e.g., Samuelson and Zeckhauser 1988, Johnson et al. 1993).

Policy *acceptance* is just as important. Policies must be accepted as just and legitimate in order for them to succeed (Tyler and Rasinski 1991), at least when stakeholders have a voice in whether and how policies are implemented (as is the case with many firms, governments, and other organizations). But what affects policy acceptance? The policy-making process seems especially important in this regard: perceived process fairness strongly affects policy acceptance (e.g., Arvai

2003, Arvai et al. 2001, McDaniels et al. 1999). Of course, people also care about outcomes; the costs and benefits of a policy to the self and to society in general affect the degree to which a person deems a policy acceptable (e.g., Sears and Funk 1991). However, as we will demonstrate, in many cases the same policy is evaluated very differently depending on how its costs and benefits are described (see Tversky and Kahneman 1981). Specifically, we find strong framing effects on policy evaluations depending on whether the description focuses on the costs for one group or the benefits for another, complementary group. In some situations, participants deem policies that are described as *disadvantaging* one group acceptable but those described as *advantaging* the complementary group unacceptable. In other situations we find the opposite; the policy described as providing gains to one group is accepted whereas the same policy described as costing the complementary group is not.

What distinguishes these situations is the type of behavior that the policy is targeting—specifically, whether the target behavior is perceived as *obligatory* or *voluntary*. Obligations are guided by the law, strong rules, and social norms. For example, we are obliged

to pay taxes, to not murder, and to financially support our children. Obligations are things one must do (or abstain from doing), and we expect others to adhere to them. Because of this, failure to comply with an obligation is generally perceived as undesirable, whereas adherence to obligations feels mostly neutral (De Bruin and Van Lange 2000; Fiske 1980; Skowronski and Carlston 1987, 1989). In the moral domain, these compulsory behaviors are often called duties (see, for example, Brandt 1964). But obligations can also exist outside of the moral domain—for example, the obligation to show up for one's scheduled shift at work.

We contrast obligations with other behaviors that are optional—think of donating to charity or of giving up one's seat to an elderly passenger on the subway. One problem with talking about these kinds of positive-but-optional behaviors is that there is no convenient term to describe them. In moral philosophy, they are most commonly called “supererogatory” (Chisholm 1963), but for our purposes, this term has two problems: first, it is technical, and second, not all the behaviors we concern ourselves with are moral. For example, it seems odd to describe maintaining a healthy weight as moral, but most people would agree that it is desirable but not required. We therefore use the term “voluntary” to describe these sorts of behaviors. Voluntary behaviors are perceived to be desirable but are not necessarily expected; they are things a person *can* do but does not *have* to do. Giving to charity is desirable, as is willingness to give up one's seat to a tired fellow passenger, but the failure to do these things would generally not warrant reproach. In other words, the absence of a voluntary behavior is generally perceived to be neutral, whereas the presence of one is desirable; i.e., it is better than expected. Like obligations, voluntary behaviors can be actions or inactions.

Because voluntary behaviors are seen as desirable but not required, whereas obligations are seen as mandatory, we hypothesize that people will apply different standards when evaluating policies targeting voluntary behaviors versus obligations. Specifically, we hypothesize that when policies target voluntary behavior, people will prefer them when they are framed as *advantaging* those who act positively rather than *disadvantaging* those who fail to do so. Conversely, when policies target obligations, we hypothesize that people will prefer them when they are framed as *disadvantaging* those who fail to fulfill obligations rather than *advantaging* those who fulfill them.

Of course, in many circumstances advantaging and disadvantaging policies do not produce identical outcomes, and in these cases, it is perfectly reasonable to prefer one policy over another. Cutting pay

for employees who do not meet productivity targets means that employees make (on average) less, whereas raising pay for employees who do meet the targets means that employees make (on average) more. In such cases, it would not be surprising if policy evaluations differed. However, many policies deal with the distribution of fixed-pie resources, where outcomes are by definition zero-sum. In these cases, advantaging and disadvantaging policies can be logically equivalent in that they lead to identical outcomes. For example, suppose that there is a limited pool of bonus money that must be distributed among employees. In this case, increasing the bonuses of some employees necessarily entails decreasing the bonuses of others. A policy that increases the bonuses of employees who meet productivity targets (advantaging productive employees) can be logically equivalent to one cutting the bonuses of those who do not (disadvantaging unproductive employees). There are many other salient examples of policies affecting the distribution of fixed-pie resources. Elite universities can only admit a limited number of students, organ banks must allocate a limited number of transplantable organs, and municipalities have a limited number of subsidized low-income rental apartments. In these cases—where making outcomes better for some necessarily makes them worse for others—the same policy can be framed as advantaging one group or disadvantaging another. We hypothesize that even when outcomes for those affected are identical under the advantaging and disadvantaging framings, people will prefer advantaging policies for voluntary behaviors and disadvantaging policies for obligations.

Why might this be the case? J. O. Urmson's feeling that “there is nothing whatever objectionable” in pressing someone to fulfill an obligation, but that there is “something horrifying” in pressing someone to act voluntarily, suggests that people may be willing to support coercive policies in the domain of obligations but not in the domain of voluntary behaviors. To see whether people generally endorse this notion, we asked people to rate (on seven-point Likert scales anchored by “totally unacceptable” and “totally acceptable”) how “morally acceptable” it was for authorities to “use the threats of punishment” to “get people to do something desirable and required (for example, following laws)” and something that was “desirable, but voluntary (for example, donating to charity).” Consistent with Urmson's intuition, respondents rated coercion as much more acceptable in the former case ( $M = 4.29$ ,  $SD = 1.70$ ) than the latter ( $M = 2.35$ ,  $SD = 1.59$ ; paired  $t(48) = 6.27$ ,  $p < 0.001$ ). This idea is further supported by findings that people believe prescriptive norms (“one should”) are more discretionary than proscriptive norms (“one should not”; Janoff-Bulman et al. 2009). (Note, however, that

prescriptive and proscriptive norms do not overlap perfectly with voluntary and obligatory behaviors. For example, one should not commit murder but should pay taxes. Both behaviors are obligatory, but the former is proscriptive whereas the latter is a prescriptive norm.)

Thus, it may be that people naturally see disadvantaging policies as (coercive) punishments and advantaging policies as (noncoercive) rewards—and therefore see a disadvantaging policy as more appropriate for obligations and an advantaging policy as more appropriate for voluntary behaviors, even when the outcomes of both policies are identical. We call this the *matching effect* in policy evaluations. There is some support for the latter half of this proposition in the literature. In a study by Tannenbaum et al. (2013), people read about a company choosing between two healthcare plans. One plan costs \$2,500, but with a \$500 discount for healthy-weight employees; the other costs \$2,000, but with a \$500 surcharge for overweight employees.<sup>1</sup> When the policy was described as an “overweight surcharge,” people saw it as more punitive and evaluated it more negatively. Tannenbaum et al. (2013) also found that evaluations of the overweight surcharge policy correlated with inferences about the policy makers’ attitudes. When the company was described as choosing the overweight surcharge plan, participants inferred that the company felt more negatively about overweight employees, and these inferences about the company’s feelings toward the overweight mediated the effect of discount/surcharge framing on policy evaluation.

Because Tannenbaum et al. (2013) posted their data online, we were able to test whether the degree to which participants saw the policy as punitive correlated with policy acceptance, even controlling for inferences about the company’s attitudes toward the overweight. Supporting the matching account, we observed a strong negative relationship between perceptions of punitiveness and policy acceptability even when holding the policy maker’s inferred attitudes constant; partial  $r(60) = -0.61$ . The Tannenbaum et al. data therefore provided initial evidence for our idea that in the realm of voluntary behavior, policies framed as advantaging those acting positively are regarded as more acceptable than identical policies framed as disadvantaging those not acting positively. Furthermore, it appears that this framing effect cannot be completely explained by the attitude inferences account advanced by Tannenbaum et al., as there is a

substantial relationship between perceived punitiveness and policy acceptance even when controlling for inferred attitudes.

In what follows, we present seven sets of studies (13 studies total) exploring the matching effect in policy evaluations. In the first six studies, we test whether a matching effect indeed emerges in judgments of policies with identical effects. In Studies 1a–1c, we find that in the realm of voluntary behaviors, people evaluate a policy as more acceptable and moral when it is framed as advantaging those acting positively rather than disadvantaging those *not* acting positively. For obligations we find the opposite pattern: people evaluate a policy framed as disadvantaging those who do not observe obligations more favorably than the same policy framed as advantaging those who do (Studies 2a and 2b). In Study 2c, we find that orthogonally varying whether the target behavior is voluntary or obligatory and the framing of the policy as advantaging or disadvantaging reveals the predicted matching effect.

We subsequently investigate the underlying causes of the matching effect. In Studies 3–5, we first test some common mechanisms underlying many framing effects. First, even though we keep outcomes identical across frames, it is possible that those judging the policy do not realize this equivalency. This possibility is investigated and ruled out in Studies 3a and 3b. Second, in Study 4, we examine a “leakage of information” explanation (Sher and McKenzie 2006) by testing whether the differences in evaluation can be explained by inferences about policy maker attitudes or intentions. The results reveal that leakage of information cannot explain the matching effect. Finally, in Studies 5a and 5b, we test whether the matching effect is moderated by (a) whether the policy targets a minority or majority or (b) whether the policy targets actions or inactions. We find that neither moderates the effect.

In Study 6, we test whether differences in policy evaluations are instead driven by participants’ perception of the policy as a form of punishment or reward. To examine this possibility, we cross framing (advantaging versus disadvantaging) with whether the policy is explicitly described as a reward or punishment, and we find that describing the policy as a reward or punishment dominates evaluations.

Finally, we conclude with a field study (Study 7) in which we test the effect of policy framing in a situation where acceptance of the policy has direct and sizeable consequences for those evaluating it.

In all studies, we predetermined sample size,<sup>2</sup> report all dependent variables and manipulations, and

<sup>1</sup> Two additional conditions were included in the study where the policy described as costing extra for overweight people had a slightly lower cost than the one framed as giving a discount to healthy-weight employees.

<sup>2</sup> Generally speaking, for all studies our planned sample size was 50 per cell for the lab studies and somewhat higher (+20%) for the Amazon Mechanical Turk studies since the latter’s data can

exclude no participants unless stated otherwise. All data sets can be retrieved from <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/C9JH4C>. Additional information about the studies can be found in the online appendix (available as supplemental material at <https://doi.org/10.1287/mnsc.2016.2539>), and we encourage readers to consult these for a full understanding of the results.

## Study 1

In Study 1 we examined whether people judge policies with identical outcomes differently when they are framed as advantaging those who act desirably versus disadvantaging those who do not. We tested this hypothesis using three different scenarios (between subjects). For brevity, only the first scenario is described fully; complete descriptions of the others are provided in Online Appendix A.

### Method

All three scenarios described a ranking system where the top “*x*” people in the ranking would receive some benefit. In these situations, moving one group up on the ranking versus moving the others down leads to exactly the same rank order, and thus to identical outcomes.<sup>3</sup> In the first scenario, participants read about a proposed government policy to increase the number of organ donors.<sup>4</sup> They read that organs are typically allocated based on how long someone has been on the waiting list, with those who have waited longest being first in line to receive an organ. Subsequently, participants in the *advantage* condition read that under the proposed policy, to reward organ donors, people who were themselves donors would be given an “extra month” on the waiting list if they ever needed an organ themselves. This was then further explained with an example (translated from Dutch): “For example, an organ donor who had been on the waiting list for three months would be treated as though he or she had been waiting for four months.” In the *disadvantage* condition, participants read that under this policy, to punish nondonors, they would receive a “one month

reduction” on the waiting list if they ever needed an organ themselves: “For example, a nondonor who had been on the waiting list for three months would be treated as though he or she had been waiting for only two months.”

After reading the proposed policy, participants rated it using the following items: “How unacceptable/acceptable is this plan?” (–5 being very unacceptable to 5 being very acceptable), “How immoral/moral is this plan?” (–5 being very immoral to 5 being very moral), and “How negative/positive is this plan?” (–5 being very negative to 5 being very positive).

The other scenarios shared the same basic structure and the same dependent variables. However, unlike Scenario 1, we did not provide explicit information about whether the policy was intended to reward or punish. Scenario 2 described a policy where people who did voluntary community service were moved up on the waiting list for a desirable apartment complex (*advantage* condition), whereas those who did not do community service were moved down (*disadvantage* condition). Scenario 3 described a company policy where employees who worked voluntary overtime would be moved up on the list of candidates for employee of the month (*advantage* condition), whereas those who did not work voluntary overtime would be moved down (*disadvantage* condition).

Participants for the first scenario were 100 undergraduates of Tilburg University (54 females, 44 males, 2 missing,  $M_{\text{age}} = 20.31$ ,  $SD = 2.45$ ) who were approached in the cafeteria or lab and asked to fill in a short questionnaire (in this and subsequent studies, materials given to the undergraduates were in Dutch, the participants’ native language). Participants for the remaining scenarios were recruited using Amazon.com’s Mechanical Turk (MTurk), an online labor pool (Scenario 2:  $N = 103$ , 45 females, 57 males, 1 missing,  $M_{\text{age}} = 34.13$ ,  $SD = 11.96$ ; Scenario 3:  $N = 101$ , 36 females, 64 males, 1 missing,  $M_{\text{age}} = 31.96$ ,  $SD = 11.54$ ).

### Results

The three policy evaluation items formed a reliable scale for all three scenarios (Scenario 1,  $\alpha = 0.90$ ; Scenario 2,  $\alpha = 0.92$ ; Scenario 3,  $\alpha = 0.95$ ) so we standardized and averaged them to create composite policy evaluation scores (results for the individual dependent variables for this study as well as the subsequent studies are reported in the online appendix). Consistent with our account, participants who read about the policy framed as advantaging organ donors rated it much more favorably ( $M = 1.09$ ,  $SD = 2.53$ ) than those who read about it framed as disadvantaging nondonors ( $M = -1.93$ ,  $SD = 2.30$ );  $t(98) = 6.26$ ,  $p < 0.001$ ,  $d = 1.25$ . Results from Scenarios 2 and 3 were similar to those of Scenario 1. Participants who read

be slightly noisier, because participants are more demographically diverse than typical college samples (Buhrmester et al. 2011). Furthermore, for all studies predicting an interaction (with the exception of Study 3), we used  $n \times 2$  since that is needed to maintain the same power (Simonsohn 2014).

<sup>3</sup> For example, consider the rank-ordered list of John, Jane, Adam, Amy, and Brad. Moving all women up one rank results in the rank order of Jane, John, Amy, Adam, and Brad. Moving all men down one rank also results in the order: Jane, John, Amy, Adam, and Brad.

<sup>4</sup> We verified that people indeed generally think that being an organ donor is positive but optional in an online pretest; 97% (86/89) thought being an organ donor was positive (versus negative), and 78% (69/89) saw it as optional (versus required).

that those doing community service would be moved up on the waiting list (Scenario 2) were more favorable toward the plan ( $M = 2.89$ ,  $SD = 1.87$ ) than participants who read that those not doing community service would be moved down ( $M = 1.06$ ,  $SD = 2.66$ );  $t(101) = 4.05$ ,  $p = 0.02$ ,  $d = 0.81$ . For Scenario 3, participants who read that those working extra voluntary overtime would be more likely to win employee of the month were more favorable toward the plan ( $M = 0.03$ ,  $SD = 3.02$ ) than those who read that workers not working overtime would be less likely to win ( $M = -1.24$ ,  $SD = 2.60$ );  $t(99) = 2.26$ ,  $p = 0.03$ ,  $d = 0.46$ .

## Study 2

Study 1 showed that people judge policies with identical outcomes differently depending on whether they are framed as advantaging one group or disadvantaging another. Policies aimed at increasing voluntary behavior are viewed more favorably (i.e., seen as more acceptable, moral, and positive) when they are framed as advantaging those who engage in voluntary behavior rather than as penalizing those who do not.

We predicted the opposite pattern for policies concerning obligations. In such cases, policies should be judged more favorably when they are framed as disadvantaging those who fail to meet obligations rather than advantaging those who do. In Study 2, we turn to this second prediction. In Studies 2a and 2b, we describe a policy designed to address a violation of obligation (parking in handicapped spaces by non-handicapped drivers in Study 2a, and cheating on a test in Study 2b). In Study 2c, we orthogonally varied the nature of the behavior (obligation versus voluntary) and the framing of the policy (advantage versus disadvantage).

## Method

Participants in Study 2a were 102 undergraduates from a Dutch university (70 females, 31 males, 1 missing,  $M_{\text{age}} = 21.27$ ,  $SD = 2.85$ ) who were randomly assigned to one of two conditions. All participants read that the government aimed to reduce parking in handicapped spots by able-bodied drivers.<sup>5</sup> In the *disadvantage* condition, participants read that able-bodied drivers caught parking in a handicapped spot would be penalized three months on the waiting list for a parking permit (in many Dutch cities, parking permits are very scarce and are allocated to wait-listed residents as they become available). In

the *advantage* condition, participants read that people who were *not* caught parking on a handicapped spot would receive a three-month bonus on parking permit waiting list (i.e., they would be treated as if they had been waiting for three extra months). After reading the scenario, participants completed the same policy evaluation measures used in Study 1.

Participants in Study 2b were 101 MTurk workers (50 females, 51 males,  $M_{\text{age}} = 35.5$ ,  $SD = 11.8$ ) who were randomly assigned to one of two conditions. All participants read that in the Netherlands, more prospective students are interested in studying medicine than can be accepted. Therefore all prospective students get ranked based on their final exam scores, with the top 1,000 being accepted. In the *advantage* condition, participants read a proposal in which students who were not caught cheating on their exam got their final score doubled. In the *disadvantage* condition, participants read a proposal in which students who were caught cheating on their exam got their final score halved. Participants then completed the same policy evaluation measures used in the previous studies.

Participants in Study 2c were 523 MTurk workers (188 females, 333 males, 2 missing,  $M_{\text{age}} = 30.27$ ,  $SD = 9.28$ ) who were randomly assigned to one of four conditions of a  $2 \times 2$  between-subjects design. All participants read that students at W. H. Taft Middle School were required to do 12 hours of community service per year. In the *obligation* conditions, participants read either that students who did not do the obligatory 12 hours of community service would be less likely to be accepted to popular wait-listed courses (*obligation-disadvantage*) or that students who completed the required 12 hours would be more likely to be accepted (*obligation-advantage*). In the *voluntary* conditions, participants read either that students who did *more* than the required 12 hours would be more likely to be accepted to these courses (*voluntary-advantage*) or that those who only did the required 12 hours would be less likely to be accepted (*voluntary-disadvantage*). Participants then completed the same policy evaluation measures used in the previous studies.

## Results

Results of these studies are consistent with our account. For obligatory behaviors, people were more favorable toward policies framed as disadvantaging those who failed to fulfill obligations rather than advantaging those who fulfilled them. For voluntary behaviors, we replicated the pattern of results from Study 1: People regarded the policy more favorably when it was framed as advantaging those who engaged in voluntary behavior rather than as penalizing those who did not.

<sup>5</sup> In a pretest, we verified that people regarded parking in handicapped spots as negative (instead of positive; 88/89, 99%) and thought that not doing so should be compulsory (instead of voluntary; 80/89, 90%).

**Study 2a.** The three policy evaluation items formed a reliable scale ( $\alpha = 0.91$ ), so we standardized and averaged them to create a composite policy evaluation score. Consistent with our expectations, participants in the disadvantage condition, who read that drivers caught parking in a handicapped spot would be moved down on the waiting list for a parking permit, were more favorable toward the plan ( $M = 1.18$ ,  $SD = 2.38$ ) than those in the advantage condition, who read that drivers *not* caught parking in a handicapped spot would be moved up ( $M = -0.14$ ,  $SD = 2.31$ );  $t(100) = 2.84$ ,  $p = 0.01$ ,  $d = 0.56$ .

**Study 2b.** The three policy evaluation items formed a reliable scale ( $\alpha = 0.98$ ), so we standardized and averaged them to create a composite policy evaluation score. Consistent with our expectations, participants in the disadvantage condition, who read students got cheating got their final score halved, were more favorable toward the plan ( $M = 2.94$ ,  $SD = 2.23$ ) than those in the advantage condition, who read that students *not* caught cheating would have their scores doubled ( $M = -0.54$ ,  $SD = 3.44$ );  $t(99) = 6.04$ ,  $p < 0.001$ ,  $d = 1.22$ .

**Study 2c.** The three policy evaluation items formed a reliable scale ( $\alpha = 0.89$ ), so we standardized and averaged them to create a composite policy evaluation score. As in Study 1, participants who judged the policy targeting voluntary behavior ([not] doing more than the required 12 hours of community service) approved of the policy more when it was framed as advantaging those doing extra hours ( $M = 0.36$ ,  $SD = 2.80$ ) rather than as disadvantaging those not doing the extra hours ( $M = -0.26$ ,  $SD = 2.56$ );  $t(256) = 1.88$ ,  $p = 0.06$ ,  $d = 0.23$ . However, participants who judged the policy targeting an obligation ([not] doing the required 12 hours) showed the opposite effect. In this case, participants favored the policy more when it was framed as disadvantaging those not doing enough hours ( $M = 1.35$ ,  $SD = 2.63$ ) rather than as rewarding those who did enough hours ( $M = 0.72$ ,  $SD = 2.76$ );  $t(263) = 1.88$ ,  $p = 0.06$ ,  $d = -0.23$ . This pattern produced a significant target behavior  $\times$  framing interaction ( $F(1, 519) = 7.1$ ,  $p = 0.01$ ,  $\eta_p^2 = 0.01$ ). We also found a main effect of target behavior such that participants favored policies targeting failures to meet obligations more than policies targeting voluntary behavior ( $F(1, 519) = 17.5$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.03$ ). As people generally care more about negative than positive behavior (see, for example, Baumeister et al. 2001), this is not surprising.

### Study 3

The preceding studies have demonstrated a matching effect in policy evaluations: when policies are designed to encourage voluntary behavior, people prefer

policies framed as advantaging rather than disadvantaging. The opposite is true for policies targeting obligations; in these cases, policies framed as disadvantaging are preferred. In the following three studies, we investigate the psychology behind these differing policy evaluations.

Studies 3a and 3b test one possible explanation for the matching effect—that people do not understand the full implications of the policy (i.e., that advantaging one person logically entails disadvantaging another). Even though the policy descriptions gave participants this information, it is possible that lack of effort, ability, or both prevented them from using it. If this were the case, increasing effort (by asking people to deliberate) or ability (by presenting policies such that their full implications are highly salient) should attenuate the matching effect. In Study 3a, we forced participants to deliberate by asking them to describe the consequences of the policy for both advantaged and disadvantaged groups. In Study 3b, we asked participants to evaluate both policies at the same time (i.e., joint evaluation; see, e.g., Hsee et al. 1999).

### Method

Participants in Study 3a were 243 MTurk workers (98 females, 145 males,  $M_{\text{age}} = 28.83$ ,  $SD = 8.08$ ) who were randomly assigned to one condition of a  $2 \times 2$  between-subjects design. Two of the conditions (the control conditions) were identical to Study 1 (but in English), in which participants read about a policy that was framed as either advantaging organ donors or disadvantaging nondonors. After reading the scenario, participants completed the same policy evaluation measure used in the previous studies. In the other two conditions (the *deliberation* conditions), participants read the same scenarios but were subsequently asked to think about the policy and its effects on nondonors as well as donors. They were then asked to describe (in a few sentences) what this policy would mean for donors, followed by the same question for nondonors. Finally, they completed the policy evaluation measures. If the results of the previous studies depended on participants not fully considering the policy's consequences for both groups, this deliberation manipulation should attenuate differences in policy evaluation.

Participants in Study 3b were 196 students from a Dutch university (129 females, 42 males, 25 missing,  $M_{\text{age}} = 19.86$ ,  $SD = 2.06$ )<sup>6</sup> who were randomly assigned to one of five conditions. Two of the conditions (the control conditions) were identical to Study 1, where participants read about a policy that was framed as either advantaging organ donors or

<sup>6</sup>Students who participated in Study 1 were excluded from participating.

disadvantaging nondonors. As in Study 3a, we also included two *deliberation* conditions in which participants were asked to deliberate about the policy and write down their thoughts about it—although unlike in Study 3a, participants were not explicitly instructed to think about the outcomes for both donors and nondonors. Finally, in the last (*joint evaluation*) condition, both policies were portrayed side by side (counterbalanced), and the evaluation questions alternated between the two policies.<sup>7</sup> After reading (and, in the deliberation conditions, writing about) the policy description(s), participants completed the same attitude measure as in the other scenarios.

## Results

**Study 3a.** The three policy evaluation items formed a reliable scale ( $\alpha = 0.96$ ), so we standardized and averaged them to create a composite policy evaluation score. In the control conditions, we replicated the pattern found in previous studies; participants who read the advantaging framing were more favorable toward the policy ( $M = 1.78$ ,  $SD = 2.42$ ) than were those who read the disadvantaging framing ( $M = -1.64$ ,  $SD = 3.05$ ). We found the same pattern in the deliberation conditions; participants who read the advantaging framing were more favorable toward the policy ( $M = 1.58$ ,  $SD = 2.84$ ) than were those who read the disadvantaging framing ( $M = -2.20$ ,  $SD = 2.77$ ). An analysis of variance (ANOVA) with framing and deliberation as factors revealed only the expected main effect of framing ( $F(1, 239) = 101.9$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.30$ ). There was no main effect of deliberation ( $F(1, 239) = 1.1$ ,  $p = 0.29$ ) nor an interaction between deliberation and framing ( $F(1, 239) = 0.26$ ,  $p = 0.61$ ).

It is possible that participants did not realize that advantaging donors entailed disadvantaging nondonors (and vice versa) even when they were explicitly asked to deliberate about the consequences of the policy for both groups. We therefore had a research assistant (who was unaware of our hypotheses and design) classify the responses of participants in the deliberation conditions as mentioning the zero-sum nature of the outcome or not. In the advantage condition, the research assistant coded whether people answered the question “what do you think the consequences of this policy are for nondonors” by saying that nondonors would move down on the list. In the disadvantage condition, he coded whether people responded to “what do you think the consequences of this policy are for donors” by saying that they would move up on the list.

<sup>7</sup> Specifically, participants indicated the judged acceptability of policy A, the acceptability of policy B, the morality of policy A, the morality of policy B, the positivity of policy A, and finally, the positivity of policy B.

By these criteria, 61% of participants explicitly mentioned that the policy outcomes were zero-sum. If the policy framing effect were driven by the 39% who may not have understood the zero-sum nature of the policy, then these participants’ responses should be different from the remaining 61%. We compared the responses of these two groups using a 2 (policy framing: advantage versus disadvantage)  $\times$  2 (participant: mentioned zero-sum versus not) ANOVA. This analysis also revealed only a main effect of policy framing ( $F(1, 111) = 34.7$ ,  $p < 0.001$ ). There was no main effect of mentioning that outcomes were zero-sum ( $F(1, 111) = 1.2$ ,  $p = 0.28$ ) nor an interaction between mentioning that outcomes were zero-sum and framing ( $F(1, 111) = 0.4$ ,  $p = 0.52$ ). Therefore, even in the most conservative test, no effect of deliberation was found.

**Study 3b.** The three policy evaluation items formed a reliable scale ( $\alpha = 0.95$ ), so we standardized and averaged them to create a composite policy evaluation score. The answers of participants in the nondeliberation conditions were similar to those in previous studies, with participants being more favorable toward the policy when it was framed as advantaging donors ( $M = 0.88$ ,  $SD = 2.61$ ) rather than disadvantaging nondonors ( $M = -2.71$ ,  $SD = 2.09$ );  $t(51) = 5.43$ ,  $p < 0.001$ ,  $d = 1.53$ . In the deliberation conditions, participants still evaluated the policy much more favorably when it was framed as advantaging donors ( $M = -0.18$ ,  $SD = 3.28$ ) rather than disadvantaging nondonors ( $M = -2.03$ ,  $SD = 2.39$ );  $t(90) = 3.04$ ,  $p < 0.01$ ,  $d = 0.65$ . An ANOVA with advantage versus disadvantage as one factor and no-deliberation versus deliberation as the second factor revealed a main effect of framing ( $F(1, 137) = 32.8$ ,  $p < 0.001$ ), no main effect of deliberation ( $F(1, 137) = 0.2$ ,  $p = 0.69$ ), and a marginally significant interaction effect ( $F(1, 137) = 3.3$ ,  $p = 0.07$ ).<sup>8</sup> The same pattern also appeared for the joint-evaluation condition, with participants being more favorable toward the policy moving donors up ( $M = 1.08$ ,  $SD = 2.82$ ) compared with the policy moving nondonors down ( $M = -3.24$ ,  $SD = 1.95$ );  $t(54) = 10.42$ ,  $p < 0.01$ ,  $d = 1.81$ . Evaluating both policies side by side thus did not eliminate the effect of framing on policy evaluations; if anything, the effect of framing was slightly larger under joint evaluation.

## Discussion

Taken together, the results of Studies 3a and 3b suggest that the differences in evaluations of the differently framed policies are not the result of participants

<sup>8</sup> In this study, there is a marginally significant reduction in the size of the matching effect. Since the size of the matching effect under deliberation is still very large ( $d = 0.65$ ,  $p < 0.001$ ), and since deliberation actually increased the matching effect in Study 3a, we believe that this marginally significant effect is likely spurious.



not understanding the equivalency between the different frames. Most participants realized that a decrease in waiting time for donors means that non-donors will have to wait longer, and vice versa. Still, they were fairly neutral about the policy advantaging donors but quite negative about the policy disadvantaging nondonors.

## Study 4

The results of Studies 3a and 3b suggest that it is unlikely that the observed differences in policy evaluations are attributable to respondents not understanding the full implications of the policies. In Study 4, we test another explanation for the matching effect: it may be that people infer attitudes of the policy maker from its framing. Such inferences would be consistent with research showing that people (correctly) believe that a communicator's choice of frame "leaks" information about his or her beliefs (McKenzie et al. 2006). For example, in Tannenbaum et al. (2013, Study 1), participants who read about a policy described as adding an extra charge for overweight employees inferred that the policy makers had more negative beliefs about the overweight compared with participants reading about a policy that was described as providing a discount for healthy-weight employees. These inferences appear to be accurate, since in Study 2 Tannenbaum et al. find that participants who hold more negative attitudes toward the overweight are more likely to choose the overweight-surcharge policy over one providing a discount for healthy-weight employees. Thus, even though the policies objectively carried the same information, participants accurately inferred the difference in attitudes between policy makers from them.

It may be that in the current studies, people understood perfectly well that a policy that advantages one group necessarily disadvantages others. They may nonetheless have responded to the choice of frame used by the policy makers. After all, even if both descriptions of the policy led to the same outcomes, there must have been a reason why the policy makers chose one description over the other. Participants may have believed that a policy framed as disadvantaging one group was implemented by the policy makers as a result of negative attitudes toward that group. If participants disagreed with these inferred attitudes, they may have expressed this by judging the policy itself more negatively. This possibility would be consistent with the Tannenbaum et al. (2013) finding that people perceived a company choosing overweight-surcharge healthcare plans as holding more negative thoughts about overweight employees compared with a company choosing healthy-weight discount plans.

This leads to a straightforward prediction. When participants are confronted with two identical policies

that are framed differently (advantaging versus disadvantaging) and read that a policy maker *chose* one description over the other, the matching effect should emerge. However, when participants read about the exact same policies but are told that the policy makers decided which description to use by a coin toss, the effect should vanish. This is what we tested in Study 4.

## Method

Participants in Study 4 were 437 workers on Amazon's Mechanical Turk (169 females, 245 males, 23 missing,  $M_{\text{age}} = 31.9$ ,  $SD = 11.3$ ) who were randomly assigned to one condition of a  $2$  (advantage versus disadvantage)  $\times 2$  (choice versus random) design. In all conditions, participants read that Luxemburg was looking into ways to increase the number of organ donors. They then read that policy makers came up with a policy under which the wait time for an organ would be influenced by whether the person in need of an organ was a donor themselves. The policy makers realize that they could present this policy in two equivalent ways—either by moving donors up on the list or by moving nondonors down on the list. It was then explained why both descriptions were equivalent and that, under both descriptions, those affected would end up with exactly the same outcomes. Participants in the *choice* conditions then read that the policy makers realized they had to choose a description and that they chose the description in which donors were moved up on the list (choice-advantage) or the description in which nondonors were moved down on the list (choice-disadvantage). Participants in the *coin flip* conditions read that the policy makers decided to flip a coin to determine which description would be used. In the random-advantage condition, they read that the coin landed heads, and therefore the advantaging description was used. In the random-disadvantage condition, they read the coin landed tails, and therefore the disadvantaging description was used. Then all participants evaluated the policy on the three questions used in Studies 1–3.

## Results

The three policy evaluation items formed a reliable scale ( $\alpha = 0.95$ ), so we standardized and averaged them to create a composite policy evaluation score. In the choice conditions, we replicated the matching effect. Participants in the advantaging condition evaluated the policy more favorably ( $M = 1.21$ ,  $SD = 2.96$ ) than did those in the disadvantaging conditions ( $M = 0.15$ ,  $SD = 3.09$ );  $t(1, 216) = 2.58$ ,  $p = 0.01$ ,  $d = 0.35$ . The same difference was found in the coin flip conditions. Participants in the coin flip-advantage condition evaluated the policy more positively ( $M = 0.93$ ,  $SD = 2.68$ ) than did those in the coin

flip-disadvantage condition ( $M = -0.44$ ,  $SD = 2.96$ );  $t(1, 217) = 3.60$ ,  $p < 0.001$ ,  $d = 0.49$ . A 2 (framing: advantage versus disadvantage)  $\times$  2 (selection procedure: choice versus coin flip) ANOVA revealed a significant main effect of framing ( $F(1, 433) = 18.81$ ,  $p < 0.001$ ). There was no main effect of selection procedure ( $F(1, 433) = 2.46$ ,  $p = 0.12$ ) and no interaction between framing and selection procedure ( $F(1, 433) = 0.32$ ,  $p = 0.57$ ). Inspection of the means shows that randomly choosing the policy frame did not attenuate the matching effect; if anything, it exacerbated it. Thus, inferences about policy maker beliefs are unlikely to explain the matching effect. It is noteworthy that even though Tannenbaum et al. (2013) found that inferred policy maker attitudes mediated differences in policy acceptance, rendering the choice of frame uninformative in this study (by describing the frame as randomly chosen) did not attenuate the matching effect.

## Study 5

Studies 1–4 reveal large differences in evaluations of a policy depending on how it is framed. They also indicate that some of the most common causes of framing effects are not responsible for this matching effect. In Study 5, we explore another potential explanation: whether the policy affects a minority or a majority of the population it addresses. Participants in the previous studies may have inferred that those engaging in voluntary behaviors or violating obligations are a minority of the population, and they may simply have preferred policies targeting smaller groups. There are two reasons they might do so. The first is economic: it may be more cost effective to move a small group up or down rather than move a large group up or down, even when the outcomes are identical. Second, people may incorrectly believe that the policies we described would have larger consequences for smaller groups. This can actually be the case—for example, when dividing a fixed amount of money between five men and one woman, giving the woman 10 extra dollars only costs each man two dollars. However, giving the men an extra 10 dollars apiece costs the woman 50 dollars. This asymmetric impact was not a feature of the policies we described in the previous studies, and in Study 4, participants are explicitly informed about the equivalence in costs for donors and nondonors between the two frames. However, it is possible that participants still hold these beliefs and that this affects their policy evaluations. In the current study, we test this possibility systematically by varying whether the policy targets a minority or majority of the population. If the matching effect is due to a preference for targeting a minority (which is the likely inference in previous studies), then we should find a reversal in judgments when the policies target a majority of the population. This was tested in Studies 5a and 5b.

## Study 5a

### Method

Participants in Study 5a were 392 workers on Amazon's Mechanical Turk (208 females, 184 males,  $M_{\text{age}} = 35.72$ ,  $SD = 9.43^9$ ) who were randomly assigned to one of four conditions of a 2 (policy framing: advantage versus disadvantage)  $\times$  2 (policy target: majority versus minority) factorial design. We used the community service scenario also used in Study 1b, in which the policy was framed as advantaging those doing community service or disadvantaging those not doing community service. In the majority conditions, we also mentioned that about 65% of the people do voluntary community service. In the minority conditions, we added a phrase mentioning that about 35% of the population does community service. If participants prefer policies targeting a small proportion, then we should replicate the preference for the advantaging policy in the minority conditions (as in Study 1b) but find the opposite effect for the majority conditions. If participants respond to the frame of the policy rather than the size of the group it targets, then we should not find an interaction effect. We also did not predict a main effect of proportion. However, finding a main effect would not necessarily be surprising, because only a small proportion doing community service may make the policy appear more important (compared with a situation where a large proportion already does community service).

### Results

The three policy evaluation items formed a reliable scale ( $\alpha = 0.95$ ), so we averaged them to create a composite policy evaluation score. Replicating Study 1b, participants in the two minority conditions evaluated the policy more favorably when it was framed as advantaging those doing community service ( $M = 2.60$ ,  $SD = 2.23$ ) rather than disadvantaging those who do not ( $M = 1.56$ ,  $SD = 2.68$ );  $t(196) = 2.97$ ,  $p = 0.003$ ,  $d = 0.42$ . A very similar pattern was found in the majority conditions: participants evaluated the policy more positively when it was framed as advantaging ( $M = 2.71$ ,  $SD = 2.10$ ) rather than disadvantaging ( $M = 1.07$ ,  $SD = 2.83$ );  $t(192) = 4.58$ ,  $p < 0.001$ ,  $d = 0.66$ . An ANOVA including both factors only reveals a significant main effect of framing ( $F(1, 388) = 28.6$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.07$ ), no main effect of proportion ( $F(1, 388) = 0.6$ ,  $p = 0.46$ ,  $\eta_p^2 = 0.00$ ), and no interaction effect ( $F(1, 388) = 1.4$ ,  $p = 0.24$ ,  $\eta_p^2 = 0.00$ ).

### Discussion

We conducted Study 5a to test whether the previously documented differences in policy evaluations

<sup>9</sup>One participant indicated his age as 1991. We assumed that was his birth year and counted him as a 24-year-old.

could be explained by a general preference for policies targeting small versus large proportions of the population, rather than an effect of framing the policy as advantaging versus disadvantaging. Had this been the case, we should have found a reversal or attenuation of the framing effect in the majority conditions compared with the minority conditions, but we did not find an interaction effect. If anything, the difference in evaluations is slightly (but not significantly) larger in the majority conditions.

## Study 5b

In Study 5b, we test one additional alternative explanation for our results. We have argued that the matching effect depends on whether the behavior addressed is seen as voluntary or obligatory. However, if participants assumed organ donation systems are opt-in (as is the case in the United States and the Netherlands), the better-liked policy in every study was also the one targeting actions (versus inactions). It is therefore possible that our results are driven by a preference for policies targeting actions. One reason for such a preference might be that actions are usually more informative about underlying dispositions than are inactions (Heider 1958, Malle 2011). Under an opt-in organ donation system, we know that a person who is an organ donor explicitly chose to be one, whereas a nondonor might have explicitly chosen to be a nondonor or might not have gotten around to making a choice yet. Likewise, a person who has parked in a handicapped spot is (by definition) willing to do so, but a person who has *not* parked in a handicapped spot may yet do so under the right circumstances. To test this alternative explanation, we ran a modified version of the organ donation scenario in which we varied whether being an organ donor was an opt-in or opt-out decision. As in Study 5a, we simultaneously varied whether a minority (in the opt-in conditions) or majority (in the opt-out conditions) of the population was said to currently be organ donors.

## Method

Participants in Study 5b were 246 workers on Amazon's Mechanical Turk (87 females, 157 males, 2 missing,  $M_{\text{age}} = 28.93$ ,  $SD = 9.23$ ) who were randomly assigned to one of four conditions of a 2 (policy framing: advantage versus disadvantage)  $\times$  2 (default: opt-in versus opt-out) factorial design. We again used the organ donation policy scenario, in which the policy was framed either as advantaging donors or as disadvantaging nondonors. In the opt-in conditions, participants read that Luxembourg uses an opt-in organ donation system, meaning that people are not organ donors unless they explicitly choose to be. Participants also read that about 25% of the population

are currently organ donors. These conditions mirror the opt-in default of the United States and the Netherlands, so these conditions should be expected to replicate our previous studies. In the opt-out conditions, the scenario closely matched that of the opt-in conditions, except that participants read that Luxemburg uses an opt-out organ donation system, meaning that people are organ donors unless they explicitly choose not to be. Participants also read that about 75% of the population are currently organ donors.<sup>10</sup> All participants then completed the policy evaluation measures described previously.

## Results

The three policy evaluation items formed a reliable scale ( $\alpha = 0.94$ ), so we standardized and averaged them to create a composite policy evaluation score. Replicating Studies 1, 3, and 4, participants in the opt-in conditions evaluated the policy more favorably when it was framed as advantaging donors ( $M = 1.88$ ,  $SD = 2.36$ ) rather than disadvantaging nondonors ( $M = 0.09$ ,  $SD = 2.97$ );  $t(123) = 3.70$ ,  $p < 0.001$ ,  $d = 0.67$ . A very similar pattern was found in the opt-out conditions: participants evaluated the policy more positively when it was framed as advantaging donors ( $M = 1.33$ ,  $SD = 2.67$ ) rather than disadvantaging nondonors ( $M = -0.17$ ,  $SD = 2.94$ );  $t(120) = 2.94$ ,  $p < 0.01$ ,  $d = 0.53$ . A  $2 \times 2$  ANOVA revealed only a significant main effect of framing ( $F(1, 241) = 21.95$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.083$ ). There was no main effect of opt-in versus opt-out ( $F(1, 241) = 1.38$ ,  $p = 0.24$ ,  $\eta_p^2 = 0.006$ ) and no interaction ( $F(1, 241) = 0.162$ ,  $p = 0.69$ ,  $\eta_p^2 = 0.001$ ).

## Discussion

Taken together, the results of Studies 5a and 5b suggest that the framing effects we have documented are not driven by whether policies target minorities versus majorities or actions versus inactions. Evaluations of the policy only depended on whether it was described as advantaging those acting voluntarily or disadvantaging those who did not, regardless of whether the policy targeted a majority or a minority, and regardless of whether donor status was active or passive. Of course, even though we were able to separate behavior type and prevalence here, in real-world situations the two covary. Adherence to obligations is almost always more common than nonadherence, whereas many voluntary behaviors are engaged in by only a minority. Indeed, as a voluntary behavior becomes more widespread, it may over time become an obligation (Rozin 1999). So although the current studies show that policies targeting minorities are not preferred per se, in most situations they will be preferred because behavior type and frequency covary so strongly.

<sup>10</sup> We chose these percentages to roughly match the actual average proportion of organ donors under opt-in and opt-out systems.

## Study 6

People evaluate equivalent policies differently depending on how they are framed. This is not due to ignorance of outcome equivalence (Study 3), inferences about policy maker attitudes (Study 4), a preference for targeting smaller groups (Study 5a), or a preference for targeting actions rather than inactions (Study 5b). What, then, can explain these robust differences in policy evaluation? One possibility is that they result from the belief one should punish bad behavior and reward good behavior. Of course, with the exception of Study 1a, we do not explicitly describe policies as punishing or rewarding, but participants may infer this from the policy descriptions. When we punish people, we often do so by taking something desirable away or by reducing their probability of a desirable outcome. When we reward people, we often do so by giving them something desirable or by increasing their probability of a desirable outcome. Thus, it may be that participants automatically construe moving someone up on a waiting list as a reward and moving someone down as a punishment. The deliberations by participants in Study 3 suggest this may be the case. Even though the policies in Study 3 only discussed moving donors up on the waiting list, or nondonors down, participants often mentioned how *rewarding* donors was only fair or how it was unfair to *punish* people for something that should be voluntary.

This reasoning suggests that explicitly describing policies as punishing or rewarding should moderate the effects we have observed thus far. When the advantaging policy is described as rewarding and the disadvantaging policy is described as punishing (i.e., consistent with participants' natural inferences), we should replicate the matching effect. When the converse is true (i.e., the advantaging policy is described as punishing and the disadvantaging policy is described as rewarding), the matching effect should be attenuated, eliminated, or even reversed. In the current study we test this hypothesis by orthogonally manipulating policy framing (disadvantaging versus advantaging) and whether this (dis)advantage is described as a reward or punishment.

## Method

Participants in Study 6 were 629 workers on Amazon's Mechanical Turk (224 females, 401 males, 4 missing,  $M_{\text{age}} = 30.12$ ,  $SD = 9.43$ ) who were randomly assigned to one of six conditions of a 2 (policy framing: advantage versus disadvantage)  $\times$  3 (policy aim: control versus framing-consistent versus framing-inconsistent) factorial design. We again used the organ donation policy scenario, in which the policy was framed as advantaging donors or disadvantaging nondonors. In the control conditions, we did

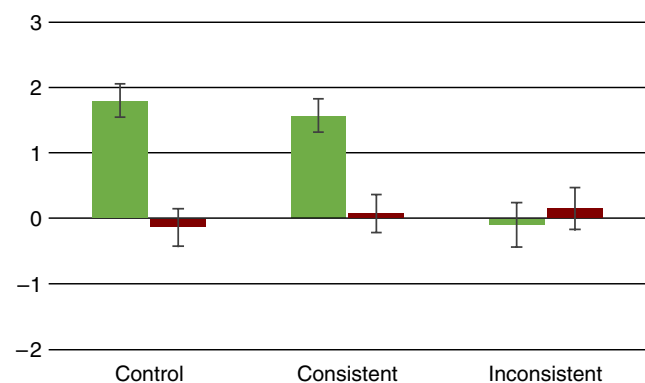
not add any further information. In the remaining conditions, we added a phrase describing the reason for the policy that was either consistent or inconsistent with what we assumed participants would infer from the policy framing. In the *framing-consistent* conditions, participants read either that to punish nondonors, they would be moved down a month on the waiting list, or that to reward donors, they would be moved up a month on the waiting list. In the *framing-inconsistent* conditions, participants read the opposite—that to punish nondonors, donors would be moved up a month on the waiting list, or that in order to reward donors, nondonors would be moved down a month on the waiting list. All participants then completed the policy evaluation measures described previously.

## Results

The three policy evaluation items formed a reliable scale ( $\alpha = 0.95$ ), so we standardized and averaged them to create a composite policy evaluation score. See Figure 1.

Replicating the previous studies, participants in the two control conditions evaluated the policy more favorably when it was framed as advantaging donors ( $M = 1.79$ ,  $SD = 2.75$ ) rather than disadvantaging nondonors ( $M = -0.13$ ,  $SD = 2.89$ );  $t(208) = 4.95$ ,  $p < 0.001$ ,  $d = 0.68$ . A very similar pattern was found in the *framing-consistent* conditions: participants evaluated the policy more positively when it was framed as advantaging ( $M = 1.57$ ,  $SD = 2.57$ ) rather than disadvantaging ( $M = 0.08$ ,  $SD = 2.87$ );  $t(207) = 3.95$ ,  $p < 0.001$ ,  $d = 0.55$ . This pattern was gone in the *framing-inconsistent* conditions, where the policy was evaluated slightly (but not significantly) more positively when it was framed as disadvantaging ( $M = 0.16$ ,  $SD = 3.21$ ) rather than advantaging ( $M = -0.09$ ,  $SD = 3.34$ );  $t(208) = -0.56$ ,  $p = 0.57$ ,

Figure 1 (Color online) Means and Errors for the Evaluation of the Policy in Study 5 That Either Provides an Advantage for Donors (Light) or a Disadvantage for Nondonors (Dark)



Note. Error bars represent standard errors of the mean.

$d = -0.03$ . This difference in the effects of framing between the framing-consistent and framing-inconsistent conditions is confirmed by a significant framing  $\times$  policy aim interaction ( $F(1, 415) = 7.2$ ,  $p = 0.01$ ,  $\eta_p^2 = 0.02$ ).

### Discussion

As hypothesized, the effect of framing on policy evaluations was eliminated when we provided information about policy aims that ran counter to what participants would naturally infer from how the policies were framed. This attenuation appears to be mostly driven by more negative evaluations of the inference-inconsistent punishing policy rather than more positive evaluations of the inference-inconsistent rewarding policy; it may be that people respond more strongly to punishments compared with rewards. The elimination of the framing effect here is especially noteworthy given that in Studies 3a, 3b, and 4, we did not attenuate the effects of framing by forcing participants to deliberate or by making the formal equivalence between the policies salient. Taken together, these results suggest that the matching effect is driven not by an insufficient understanding of the policies' implications for both groups affected but rather by participants' perception of the policy as a form of punishment or reward.

### Study 7

In the previous studies, we found that policies aimed at increasing voluntary behaviors are evaluated more favorably when they are framed as advantaging those engaging in the behavior rather than disadvantaging those who do not. In these studies, participants evaluated hypothetical policies that were not directly relevant to their own well-being. In the current field experiment, participants were asked to evaluate a policy with direct personal consequences.

We recruited our participants from a psychology course in which students have to conduct an experiment and create a poster presenting their results. Posters were graded pass/fail, but the authors of the three best posters won a €40 gift certificate. We used this naturally occurring situation of ranking-based rewards to test whether the matching effects observed in our previous studies would be obtained when the policy evaluated has direct consequences for the evaluator.

### Method

All 267 students enrolled in the Research Practicum course at Tilburg University (a second-year course in this university's psychology curriculum) were sent an email in which they were asked to complete a questionnaire. The email stated that this questionnaire concerned their evaluation of the overall grading for

the course as well as the criteria based on which the €40 gift certificates were awarded. The 82 students who completed the questionnaire<sup>11</sup> (15 males, 67 females,  $M_{\text{age}} = 20.96$ ,  $SD = 3.11$ ) were randomly assigned to one of two conditions. In both conditions, students read that to determine which posters should win the gift certificate, the posters would be judged on four criteria, each worth between 0 and 25 points. We made it clear to students that these criteria would apply only for awarding the bonus prize, not for the actual (pass/fail) grade. In the *advantage* condition, students read that, under a new proposal, there would be 20 bonus points assigned when the poster was written in English.<sup>12</sup> In the *disadvantage* condition, students read that posters not in English would get a 20-point deduction. Then, all students read that all posters would be ranked according to their total score and the three posters with the highest score would receive a €40 gift certificate.

Students read that the coordinator of the course was interested in their opinions about this policy. If they did not agree with the policy, they could vote against it by (1) indicating why they objected to it and (2) providing their email address so that they could be contacted by the course coordinator. As in the previous studies, students were asked to evaluate the proposed policy on the following items: "How unacceptable/acceptable is this plan?" (–5 being very unacceptable to 5 being very acceptable), "How immoral/moral is this plan?" (–5 being very immoral to 5 being very moral), and "How negative/positive is this plan?" (–5 being very negative to 5 being very positive). Finally, students were asked to estimate the chance that their poster would be one of the winners in the ranking (–3 being very small to 3 being very large), whether their own posters and supervisors are Dutch or English, and whether the policy makers want to punish use of Dutch language versus reward use of English language (–3 being punish Dutch to 3 being reward English).<sup>13</sup>

<sup>11</sup> Note that 103 students started the questionnaire. We did not include the 21 students that started but did not complete the questionnaire in our sample.

<sup>12</sup> The students' native language is Dutch and the course is taught in Dutch, so we expected that writing a poster in English would be seen as positive but optional. We verified this in a pretest: most respondents thought writing the poster in English was positive (74/89; 83%) but should be optional (68/89; 76%).

<sup>13</sup> Of the 82 students, 13 wrote their poster in English, and 69 wrote in Dutch. Including both condition and whether one's own poster was in English or Dutch in a  $2 \times 2$  ANOVA only reveals a main effect of poster language on policy judgment such that those who made their posters in English liked both policies better ( $M = 2.49$ ,  $SD = 2.48$ ) than those who made their poster in Dutch ( $M = -2.13$ ,  $SD = 2.06$ ), which makes sense because both policies would favor posters in English. No significant interaction effect was found. We

## Results

As expected, students in the disadvantage condition were more likely to state that they objected to the proposed policy (38/43; 88%) than were students in the advantage condition (26/39; 67%);  $\chi^2(1, N = 82) = 5.62, p = 0.02$ . Most of the objections were either that the policy was not fair because it was not announced before the course started that posters in English would have a better chance to win, or that it seemed unfair to punish students for writing posters in Dutch. Again, the three policy evaluation items showed high reliability ( $\alpha = 0.93$ ) and were combined into one composite measure. Participants in the advantage condition were more favorable toward the plan ( $M = -0.31, SD = 2.86$ ) than were those in the disadvantage condition ( $M = -2.39, SD = 2.12$ );  $t(80) = 3.74, p < 0.001, d = 0.84$ .

To test whether these differences in policy evaluations could be explained by inferences about the policy maker intentions (see Study 4), we tested whether these inferences mediated the policy judgments. Participants in both conditions believed, on average, that the policy maker intended to reward English rather than punish Dutch, but those in the disadvantage condition less so ( $M = 4.88, SD = 2.23$ ) than participants in the advantage condition ( $M = 5.72, SD = 1.31$ );  $t(80) = 2.04, p = 0.05$ . Furthermore, perceived intentions correlated with policy evaluations ( $b = 0.35, SE = 0.15, p = 0.02$ ) and partially mediated the direct path from condition to policy evaluations ( $b = 0.29, SE = 0.19, p = 0.02$  ( $0.03 < \text{confidence interval} < 0.80$ )). On the basis of the direct path (in the full model) ( $b = 1.79, SE = 0.55, p < 0.01$ ), inferences about the policy makers' intentions thus explained 10.6% of the effect of condition on policy evaluations. Thus, although inferences do appear to play a role in policy judgments, they cannot fully explain the matching effect.

## Discussion

These results show that policy framing can influence evaluations of policies that directly affect the evaluator. In the current study, students who would be personally affected by a proposed policy evaluated this policy more positively when it was framed as advantaging English posters rather than disadvantaging Dutch posters. This is especially remarkable given that most students would have been made relatively worse off—only 13 of 82 posters (16%) were in English. (The experiment was intentionally conducted after students had turned in their posters to be sure that the experiment would not affect the students'

find a similar pattern of results in a replication of Study 1a, where we also measure whether participants were organ donors themselves. Thus, whereas personal relevance does influence judgments of policies targeting this behavior in general, it does not affect the matching effect.

work for the course.) Naturally, students considered the policy somewhat unfair in both conditions, with many complaining that they would have created the poster in English if they knew about this possible rule change in advance. These objections illustrate that students in both conditions realized the negative consequences for posters in Dutch. Nonetheless, robust differences between conditions in policy evaluation and support emerged.

## General Discussion

Thirteen studies reveal that the way a policy is framed dramatically influences the way people evaluate it. These differences in judgment are consistent with a “matching effect”: When the policy addresses voluntary behaviors, policies framed as advantaging those who engage in the behavior are judged more favorably than policies framed as disadvantaging those who do not. However, when the policy addresses obligations, this pattern flips; in this case people prefer policies when they are framed as disadvantaging those who fail to fulfill obligations rather than advantaging those who fulfill them. These effects are the result of people judging a policy not only based on its outcomes (which were identical across the different frames) but also based on perceptions of the policy as a reward or punishment. Of course, these studies do not explain why people prefer to punish violations of obligations and reward compliance with voluntary norms. While this broad question lies outside the scope of this paper, we believe this is likely due to reciprocity motives. Violations of group norms and rules lead to costs for other group members, and based on principles of reciprocity, costs can then be imposed on the perpetrator (Guala 2012). On the other hand, adherence to voluntary behaviors typically benefits other group members, so here, reciprocity implies that those behaviors should be rewarded.

The matching effect is surprisingly large given that we only change the framing of the policy, not its outcomes. Often, changes in framing did not merely change the extremity of the policy evaluations but even their valence. In several studies, participants moved from, on average, being favorable toward the policy (a positive mean rating) to a strong dislike of the policy (a negative mean rating). For example, in Study 1a, a large majority of participants assigned a positive judgment to the policy moving donors up (72% rated it on the positive side of the scale), while only 20% assigned it a positive rating when it was described as moving nondonors down.

## Implications for Policy Makers

That the framing of a policy can have strong influences on the acceptability appears to be at odds

with common approaches to policy making. Often policy makers focus on the expected outcomes of policies, basing these expectations on the assumption that those targeted by the policies are rational actors (Dror 1983). If people are indeed rational actors, then the framing of a policy should not affect how they evaluate these policies—certainly not to the degree documented here. Instead, the current findings are in line with recent empirical work showing that people's responses to policies can often be at odds with the assumption of rational actors common with most policy makers (see, for example, Shafir 2001). For example, it has been found that defaults can strongly influence the decision to become an organ donor (Johnson and Goldstein 2003), partitioning choices has a large influence on investment allocations (Langer and Fox 2005), and adding evaluative categories improves the use of numeric information (and thus leads to better choices; see Peters et al. 2009). These findings indicate problems in the current assumptions underlying how policies are designed, but they also suggest promising ways in which the framing of policies can be used in "choice architecture" (Johnson et al. 2012), and they have clear practical implications.

For example, our results suggest that policy makers would be wise to ensure that policy framing fits the general evaluation of the behavior which is targeted by the policy. If the behavior is mostly seen as voluntary, the policy will be more likely to be accepted if it is framed as a advantaging; if, on the other hand, the behavior is seen as an obligation, the policy will be more likely to be accepted if it is framed as disadvantaging. When feasible, the policy maker could even tailor the description of the policy to the individual attitudes of the employees. For example, a company that wants to increase recycling could frame a policy as punishing those who do not recycle for those employees that believe recycling is a duty, not optional. But it could frame the policy as rewarding recycling for those who believe it to be desirable but optional.

### Implications for Other Research

Our findings are consistent with those of Tannenbaum et al. (2013), who found that people evaluated health insurance plans more positively when they were framed as rewarding healthy-weight employees rather than punishing overweight employees. However, we expand on their results by predicting and finding the opposite pattern for policy evaluations regarding obligations. Furthermore, we find that the account advanced by Tannenbaum et al. to explain their results—namely, that people infer policy maker attitudes from the policy framing, and that this drives policy evaluations—cannot fully explain the current

data. Although we conceptually replicate the inferred attitudes effect in Study 7, where these inferences partially mediated the effect of framing, a substantial direct effect of framing on policy evaluations remained. We also found that framing effects on policy evaluation emerged when the policy maker was not explicitly described as choosing one policy description over the other, and even when policy makers were described as choosing a frame randomly (see Study 4). It therefore seems that in addition to inferences about policy maker intentions, perceptions of policies as rewarding or punishing exert a large effect on policy evaluations.

The matching effect we document here may also qualify several well-known framing effects. Across several domains described below, research has shown that people respond more negatively to procedures that are framed as disadvantaging compared with procedures with the same outcome that are framed as advantaging. Generally speaking, these effects have been explained as a general aversion to disadvantaging or losses. The current results, however, suggest a boundary condition on this aversion—namely, that it should be moderated by perceptions of the target group.

For example, Kahneman et al. (1986) documented differences in the perceived fairness of different profit-seeking behaviors. In one of their studies (Study 4), they asked half of participants how fair they thought it was for a company that is making a loss to reduce wages by 7% under conditions of no inflation. The other half of participants were asked how fair it was for the company to only increase wages by 5% under conditions of 12% inflation. Even though the outcomes for workers are identical in both cases, participants judged the first (reduction by 7%) to be unfair but a pay increase of only 5% (a 7% pay cut in real terms) to be fair. Kahneman et al. (1986) conclude from these results that people generally find losses unfair but think nongains are fair. Our results agree with that theory, to the degree that the behavior being targeted is generally seen as desirable. If the company needed to save money because all their employees slacked off, left work early, or broke compulsory rules in other ways, we would expect these judgments of unfairness to disappear and people to actually prefer the wage decrease over a lack of compensation for inflation. In other words, people are not generally adverse to disadvantaging policies, but only when it does not match the behavior of those being targeted.

We pretested the behaviors used in the current studies to be sure that people largely agreed about whether they are voluntary or obligatory. What happens in situations in which it is ambiguous whether

the behavior is voluntary? Mulder (2008) investigated the effects of framing a policy as rewarding or punishing from this angle. If people indeed have an intuition that punishments should be addressing obligations while rewards should be reserved for voluntary behaviors, then for ambiguous behaviors the choice of frame should be informative about how voluntary this behavior is. This is exactly what Mulder (2008) found. In Mulder's study, participants observed a (fictitious) dictator game in which the dictators would either get a bonus if they allocated at least seven tokens to the receiver or get a penalty if they did not allocate at least seven tokens to the receiver. These participants then judged the morality of two dictators, one that followed the request (assigned 7+ tokens) and one that did not. It was found that, when dictators not giving at least seven tokens were punished (rather than those giving more than seven rewarded), participants judged these dictators as being less moral. Similarly, framing a norm as prescriptive (e.g., one should speak the truth) versus proscriptive (one should not lie) affects both how people view transgressions of these norms as well as adherence. A person telling a lie is seen as less moral when the norm is portrayed as "one should not lie" compared with when the norm is framed as "one should tell the truth." On the other hand, telling the truth is perceived to be more moral in the latter situation compared with the former (Janoff-Bulman et al. 2009). The results of Mulder (2008) and Janoff-Bulman et al. (2009) can thus be seen as consistent with the matching hypothesis in that people expect the type of behavior targeted by a policy to match the framing of the policy, with policies that punish or exclude being reserved for obligations, and failures to act as obliged thus being seen as more immoral.

At first glance, the results we have reported here appear similar to research on "regulatory fit" (Higgins 2000, Cesario et al. 2008). According to regulatory fit theory (Higgins 2005), "fit" between the manner of goal-pursuit and one's "regulatory orientation" (either chronic or primed) leads to an experience of "rightness" and increased engagement with the goal-directed activity (Cesario et al. 2008). Goals can be pursued using either an "eager" strategy (seeking advancements) or a "vigilant" strategy (avoid closing off possible advancements). The former means fit with a "promotion focus" orientation (which emphasizes achieving aspirations); the latter means fit with a "prevention focus" orientation (which emphasizes fulfilling obligations). When people experience regulatory fit, they engage more with the task, which in turn results in stronger evaluative reactions to the activity of goal pursuit and makes them feel more right about the activity (Higgins 2005). Assuming that considering obligations primes a prevention goal and

that considering voluntary behaviors primes a promotion goal, could the current results be explained by regulatory fit? There are two reasons to think not: one conceptual, one empirical.

From a conceptual standpoint, although the distinction between obligations and voluntary behavior maps readily to the distinction between prevention and promotion focus, eager means do not necessarily entail rewards, and vigilant means do not necessarily entail punishments. This is because regulatory fit theory defines means (eager versus vigilant) based on their intended *ends* ("ensure the presence of positive outcomes" or "ensure the absence of negative outcomes"; see Cesario et al. 2008, p. 445). Indeed, in one of the only papers applying regulatory fit to third-party judgments, the authors explicitly define eagerness/vigilance as *orthogonal* to reward/punishment. Both "encouraging desired behaviors (e.g., encouraging you to succeed, setting up opportunities for you to engage in rewarding activities, etc.)" and "withdrawing...support (e.g., taking away a privilege, acting disappointed when you fail, etc.)" are described as "eager" means (Camacho et al. 2003, p. 505).

Empirically, the size of the effects we report and their robustness to deliberation are incompatible with an explanation based on regulatory fit. Regulatory fit effects are generally quite small. A recent meta-analysis that does *not* control for publication bias finds an average effect size of  $r = 0.27$ , which—especially realizing this is an upper bound estimate—is much smaller than the effects documented here (Motyka et al. 2014). Finally, regulatory fit theory would have predicted effects of deliberation that we did not observe. Regulatory fit is argued to affect subsequent judgments via increased engagement with the task, which results in stronger evaluative reactions to the activity of goal pursuit and makes participants feel more right about the activity (Higgins 2005). Assuming that feelings of rightness feed into judgments of morality, if these feelings of rightness are a fluency-like effect, then deliberation should have reduced these matching effects (Lee and Aaker 2004). If feelings of rightness are the result of increased engagement with the task and greater elaboration, then deliberation should have increased these effects (Higgins 2005). We, on the other hand, find no effects of deliberation in either direction. Altogether, then, the effects documented here are not explained by regulatory fit theory as currently conceived.

### Moderators and Boundary Conditions

In several studies, we attempted to attenuate the matching effect, but deliberation, joint evaluation, and having the description of the policy randomly determined had no effect. This suggests that the matching effect is most readily moderated not by focusing on



policy descriptions, but rather by focusing on the matching effect's precursors: perceptions of the target behavior as voluntary or obligatory and perceptions of the policy as rewarding or punishing. As we showed in Study 5, policy evaluations are affected by whether the policy is explicitly described as rewarding or punishing. Other factors may affect these perceptions as well. For example, source effects may affect policy perceptions—if a group known for aggressively opposing Y proposes a policy framed as advantaging not Y, it is likely that many will still perceive this policy as punishing Y.

Since the matching effect results from a match between the perceptions of the behavior and perceptions of the policy, changes in perceptions of a behavior (from voluntary to obligatory, or vice versa) will change evaluations of policies affecting it. For example, as opposition to smoking became moralized in the United States over the last 60 years, legal restrictions on smoking gained public support (Rozin and Singh 1999). The notion that people support banning things they find immoral is intuitive, but the realization that people and groups can differ on the extent to which they believe behaviors to be voluntary or obligatory can lead to less intuitive consequences for policy support. For example, many liberals see assistance to the needy as a social obligation—and thus they support coercive policies (i.e., taxation and redistribution) to provide this assistance. Conservatives, on the other hand, believe providing assistance to the needy should be the voluntary choice of individuals, and thus oppose such policies—even though they are just as likely as liberals to voluntarily give to charities (even nonreligious charities) (Margolis and Sances 2013). Similarly, Republicans are less likely than Democrats to buy a product on which a carbon-reduction tax has been imposed, but they do not differ from Democrats when, instead of a tax, a logically identical “carbon-reduction offset” is imposed (Hardisty et al. 2010). Assuming that Democrats are more likely than Republicans to see carbon reduction as an obligation, and that taxes are perceived as more coercive than offsets, the matching hypothesis would predict that Republicans will find a carbon tax less acceptable than a carbon offset. Of course, for these particular cases, Democrats and Republicans likely differ in their valuation of policy outcomes as well. However, the matching hypothesis can explain effects of framing above and beyond outcomes, such as differences in evaluations of outcome-equivalent “taxes” and “offsets.”

## Conclusion

The 13 studies we report here support J. O. Urmson's intuitions regarding the acceptability of coercion in

different domains: policies with the same outcomes are evaluated very differently depending on whether they are framed as punishment and rewards, with punitive policies being seen as appropriate to encourage the observance of duties but inappropriate for behavior that should be voluntary. For rewarding policies, the reverse is true. This matching effect produces large effects on policy evaluations and is robust to a range of debiasing interventions.

## Supplemental Material

Supplemental material to this paper is available at <https://doi.org/10.1287/mnsc.2016.2539>.

## Acknowledgments

The authors thank Arnoud Plantinga for his help categorizing the open-ended responses in Experiments 3a and 3b. The authors thank T. Bradford Bitterly, Philippe van de Calseyde, and the Wharton Decision Processes lab for their helpful comments and suggestions.

## References

- Andreoni J, Harbaugh W, Vesterlund L (2003) The carrot or the stick: Rewards, punishments, and cooperation. *Amer. Econom. Rev.* 93(3):893–902.
- Arvai JL (2003) Using risk communication to disclose the outcome of a participatory decision-making process: Effects on the perceived acceptability of risk-policy decisions. *Risk Anal.* 23(2):281–289.
- Arvai JL, Gregory RS, McDaniel TL (2001) Testing a structured decision approach: Value-focused thinking for deliberative risk communication. *Risk Anal.* 21(6):1065–1076.
- Baumeister RF, Bratslavsky E, Finkenauer C, Vohs KD (2001) Bad is stronger than good. *Rev. General Psych.* 5(4):323–370.
- Brandt RB (1964) The concepts of obligation and duty. *Mind* 73(291):374–393.
- Buhrmester M, Kwang T, Gosling SD (2011) Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspect. Psych. Sci.* 6(1):3–5.
- Camacho CJ, Higgins ET, Luger L (2003) Moral value transfer from regulatory fit: What feels right is right and what feels wrong is wrong. *J. Personality Soc. Psych.* 84(3):498–510.
- Cesario J, Higgins ET, Scholer AA (2008) Regulatory fit and persuasion: Basic principles and remaining questions. *Soc. Personality Psych. Compass* 2:444–463.
- Chisholm R (1963) Supererogation and offence: A conceptual scheme for ethics. *Ratio* 5(1):1–14.
- Cialdini RB, Demaine LJ, Sagarin BJ, Barrett DW, Rhoads K, Winter PL (2006) Managing social norms for persuasive impact. *Soc. Influence* 1(1):3–15.
- De Bruin ENM, Van Lange PAM (2000) What people look for in others: Influences of the perceiver and the perceived on information selection. *Personality Soc. Psych. Bull.* 26(2): 206–219.
- Dror Y (1983) *Public Policy Making Reexamined*, 2nd ed. (Transaction Publishers, New Brunswick, NJ).
- Fiske ST (1980) Attention and weight in person perception: The impact of negative and extreme behavior. *J. Personality Soc. Psych.* 38(6):889–906.
- Guala F (2012) Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral Brain Sci.* 35(1):1–59.

- Hardisty DJ, Johnson EJ, Weber EU (2010) A dirty word or a dirty world? Attribute framing, political affiliation, and query theory. *Psych. Sci.* 21(1):86–92.
- Heider F (1958) *The Psychology of Interpersonal Relations* (John Wiley & Sons, New York).
- Higgins ET (2000) Making a good decision: Value from fit. *Amer. Psychologist* 55(11):1217–1230.
- Higgins ET (2005) Value from regulatory fit. *Current Directions Psych. Sci.* 14(4):209–213.
- Hsee CK, Loewenstein GF, Blount S, Bazerman MH (1999) Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psych. Bull.* 125(5):576–590.
- Janoff-Bulman R, Sheikh S, Hepp S (2009) Proscriptive vs. prescriptive morality: Two faces of moral regulation. *J. Personality Soc. Psych.* 96(3):521–537.
- Johnson EJ, Goldstein D (2003) Do defaults save lives? *Science* 302(5649):1338–1339.
- Johnson EJ, Hershey J, Meszaros J, Kunreuther H (1993) Framing, probability distortions, and insurance decisions. *J. Risk Uncertainty* 7(1):35–51.
- Johnson EJ, Shu SB, Delleart GBGC, Fox C, Goldstein DG, Haubl G, Larrick RP, et al. (2012) Beyond nudges: Tools of a choice architecture. *Marketing Lett.* 23(2):487–504.
- Kahneman D, Knetsch JL, Thaler RH (1986) Fairness and the assumptions of economics. *J. Bus.* 59(4):S285–S300.
- Langer T, Fox CR (2005) Bias in allocation among risky and uncertain prospects: Partition-dependence, unit-dependence, and procedure-dependence. Working paper, University of Muenster, Muenster, Germany.
- Lee AY, Aaker JL (2004) Bringing the frame into focus: The influence of regulatory fit on processing fluency and persuasion. *J. Personality Soc. Psych.* 86(2):205–218.
- Malle BF (2011) Attribution theories: How people make sense of behavior. Chadee D, ed. *Theories in Social Psychology* (Wiley-Blackwell, Malden, MA), 72–95.
- Margolis M, Sances M (2013) Who really gives? Partisanship and charitable giving in the United States. Working paper, University of Pennsylvania, Philadelphia. <http://ssrn.com/abstract=2148033>.
- McDaniels RS, Gregory RS, Fields D (1999) Democratizing risk management: Successful public involvement in local water management decisions. *Risk Anal.* 19(3):497–510.
- McKenzie CRM, Liersch MJ, Finkelstein SR (2006) Recommendations implicit in policy defaults. *Psych. Sci.* 17(5):414–420.
- Motyka S, Grewal D, Puccinelli NL, Roggeveen AL, Avnet T, Daryanto A, de Ruyter K, Wetzels M (2014) Regulatory fit: A meta analytic synthesis. *J. Consumer Psych.* 24(3):394–410.
- Mulder LB (2008) The difference between punishments and rewards in fostering moral concerns in social decision making. *J. Experiment. Soc. Psych.* 44(6):1436–1443.
- Peters E, Dieckmann NF, Västfjäll D, Mertz CK, Slovic P, Hibbard JH (2009) Bringing meaning to numbers: The impact of evaluative categories on decisions. *J. Experiment. Psychol.: Appl.* 15(3):213–227.
- Rozin P (1999) The process of moralization. *Psychol. Sci.* 10(3):218–221.
- Rozin P, Singh L (1999) The moralization of cigarette smoking in the United States. *J. Consumer Psych.* 8(3):321–337.
- Samuelson W, Zeckhauser R (1988) Status quo bias in decision making. *J. Risk Uncertainty* 1(1):7–59.
- Sears DO, Funk CL (1991) The role of self-interest in social and political attitudes. *Adv. Experiment. Soc. Psych.* 24:1–91.
- Shafir E, ed. (2001) *The Behavioral Foundations of Public Policy* (Princeton University Press, Princeton, NJ).
- Sher S, McKenzie CRM (2006) Information leakage from logically equivalent frames. *Cognition* 101(3):467–494.
- Simonsohn U (2014) No-way interactions. *Data Colada* (blog), March 12, <http://datacolada.org/17>. [DOI: 10.15200/winn.142559.90552.]
- Skowronski JJ, Carlston DE (1987) Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *J. Personality Soc. Psych.* 52(4):689–699.
- Skowronski JJ, Carlston DE (1989) Negativity and extremity biases in impression formation: A review of explanations. *Psych. Bull.* 105(1):131–142.
- Tannenbaum D, Valasek CJ, Knowles ED, Ditto PH (2013) Incentivizing wellness in the workplace: Sticks (not carrots) send stigmatizing signals. *Psych. Sci.* 24(8):1512–1522.
- Tversky A, Kahneman D (1981) The framing of decisions and the psychology of choice. *Science* 211(4481):453–458.
- Tyler TR, Rasinski K (1991) Procedural justice, institutional legitimacy, and the acceptance of unpopular U.S. supreme court decisions: A reply to Gibson. *Law Soc. Rev.* 25(3):621–630.
- Urmson JO (1958) Saints and heroes. Melden A, ed. *Essays in Moral Philosophy* (University of Washington Press, Seattle), 198–216.