The Policy Consequences of Information: Three Essays on Being (Un)Informed

A dissertation submitted to the
Department of Social and Decision Sciences
in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy in Behavioral Decision Research

by

David Hagmann

Dissertation Committee:

George Loewenstein (Chair)
Russell Golman
Alex Imas

Carnegie Mellon University
Pittsburgh, Pennsylvania

January 2019

Abstract

Information in economics is viewed as desirable to the extent that it leads to better decisions. Growing evidence, however, finds that people are often motivated to avoid instrumentally useful information – and sometimes are better off for doing so. I develop and validate a scale and show that the desire to obtain and avoid information is domain-specific, stable over time, and predicts consequential behavior. One consequence of such a desire to avoid information is that some behavioral interventions, e.g. energy reports comparing one's own use to that of neighbors, may have a hedonic cost that is not accounted for (Chapter 1). Providing information can also have a more direct policy cost. When tackling challenging societal problems like climate change, we may be motivated to provide information to citizens and policymakers about all available options. However, painless nudges, intended to be complementary policies, can come to be viewed as solutions on their own. They can then crowd-out support for more effective policies, ultimately undermining the policy objective (Chapter 2). Finally, behavioral interventions have been subject to the criticism that they are manipulative and would be ineffective if transparently disclosed. I show that, contrary to this fear, telling people that they were randomly assigned to a default option does not diminish the effectiveness of the nudge (Chapter 3).

Acknowledgments

I am extremely grateful for the support and guidance of my adviser, George Loewenstein, who is a co-author on all three essays in this dissertation and on many other projects. Working with him has been an incredibly enjoyable and rewarding experience and his curiosity and passion for research have inspired me throughout my time at Carnegie Mellon. I want to thank Cleotilde "Coty" Gonzalez, who has welcomed an economist into her psychology lab and offered me a new way of thinking about decisions under uncertainty, along with generous office space.

I'm grateful to Russell Golman for his advice and guidance, on this dissertation as well as on other projects. I always left his office with a better understanding of a problem and often a novel way of approaching it. I am also thankful to Alex Imas for advising me on my research, teaching an illuminating class on experimental economics that continues to motivate and inform my work, and offering invaluable professional mentorship along the way.

I'm especially grateful to Emily Ho, who is a co-author on Chapters 1 and 2. I will always remember the late evenings debating how to best organize our Rmarkdown documents. Collaborating with her has made every part of research (and conferences) more enjoyable and her dedication has inspired me to work harder when we got stuck.

I want to thank Alex Rees-Jones, who hosted me at The Wharton School for an intellectually stimulating year. Working with him has made thinking about taxes fun; and he supported me with extremely valuable mentorship during and after my visit. I also thank the Swiss National Science Foundation for providing me with a predoctoral fellowship for the year. I am also grateful to my undergraduate adviser, Janis Barry. She introduced me to the way economists think about the world and challenged my policy views.

I also want to thank the many graduate students in the Department of Social and Decision Sciences whose camaraderie has made pursuing a PhD a pleasure. My research would not have been possible without the tireless support of our lab managers and administrative assistants. I am particularly grateful to Fei Lu and Orsi Kovacs at the Dynamic Decision Making Laboratory and Benjamin Schenck at the BEDR Policy Lab. I also want to thank Mary Anne Hunter, whose cheerful presence has enriched my experience, and Rosa Stipanovic, who processed uncountable research reimbursements on my behalf.

Last, but certainly not least, I am ever grateful to my parents, Marlise and Ernst. Without their unconditional love, support, and encouragement, none of this would have been possible.

Contents

List of Tables

List of Figures

Introduction

The acquisition of information holds the promise to improve the quality of our decisions and, hence, our well-being. The more we know about the options that are available to us, the more likely we are to choose the ones that maximize our utility. This optimistic view dates back to George Stigler, whose seminal work introduced to economics the idea of information as a scarce resource: buyers in a market lacked information about the products and prices available to them and sellers paid a cost to inform them through advertisements (Stigler 1961). In the nearly 60 years since, economists have tended to adhere to the Stiglarian assumption that information is useful and desirable to the extent (and only to the extent) that it helps us make better decisions. When information is freely available and there are no strategic considerations, a decision-maker should always obtain it and never incur any costs to avoid it.

But a psychological account of information suggests that our relationship with information is more complex than that. For example, the prospect of carrying a gene that puts us at risk of an incurable disease does not just affect our future consumption choices, but also creates anxiety in the present. People may ultimately prefer ignorance, even when learning that they carry the gene would influence economically consequential decisions (e.g. whether to have children or how much to save for retirement). Oster, Shoulson, and Dorsey (2013) look at people who have one parent with Huntington's disease, and hence have a 50% chance of suffering from the disease themselves. They all had access to a free diagnostic test that would resolve the uncertainty, and those who opted to take the test and found they had the condition really did change their behavior. However, fewer than 10% of people in their sample chose to get tested and instead lived their lives like those who learned they did not carry the mutation.

While people may have many reasons not to get tested, including the monetary costs of the test, Ganguly and Tasoff (2016) conduct a laboratory experiment that eliminates all (economic) costs: participants consent to having their blood drawn and participate in a series of laboratory tasks in exchange for a fixed payment and bonus earnings. At the end of the study, participants are offered the option of forgoing part of their participation payment in exchange for *not* receiving test results for a sexually transmitted disease. Such avoidance not only shows the hedonic cost of learning unpleasant information, but also shows the potential costs to society: those who fail to learn they carry a communicable disease may then be more likely to spread it to others.

Information avoidance is not limited to the health domain. Participants in another laboratory experiment were ranked based on their attractiveness as rated by those of the opposite sex (Eil and Rao 2011). At the end of the experiment, they had the opportunity to learn what their rank was compared to others in the session. Some participants (those rated as least attractive) were willing to forgo part of their payment to avoid learning their true rank. While people may especially treasure their good health and their desirable self-image, there is also ample evidence that they avoid bad financial news: Sicherman et al. (2016), for example, find that those with brokerage accounts invested in stocks are less likely to log onto their accounts on days when markets are down and they might expect to observe losses.

Although previous work has documented substantial degrees of avoidance across domains (see Golman, Hagmann, and Loewenstein 2017 for a systematic review), these studies have been conducted across subjects. It may be that there are some people who do not wish to have any (potentially painful) information, while the majority of people behave like standard economic theory assumes and obtain instrumentally useful information. Alternatively, it could be that many people wish to avoid *some* information, but those who are afraid of learning about their health risks may be receptive to information about their finances.

Chapter 1 investigates this question empirically and develops a psychometric scale to measure information preferences across three economically consequential domains: health, finance, and self-image. The desire to obtain or avoid potentially painful information informs many high-stakes, economically relevant decisions and may be as fundamental to understanding them as are risk and time preferences. I show that information preferences are a stable construct and that the measure is predictive of consequential decisions to obtain (or avoid) information. The scale provides evidence for a high prevalence of avoidance: most respondents were willing to avoid at least some information. Moreover, while the desire to avoid information in one domain is correlated with the desire to do so in another, I show that avoidance is domain-specific. Information preferences, moreover, are correlated with both risk and time preferences: those who are more tolerant of risk and those who are more patient are also more willing to obtain information. This is consistent with a view in which information (and beliefs) enter directly into someone's utility function (see Golman, Hagmann, and Loewenstein 2017): news then resembles a risky gamble that may either return a gain (good news) or a loss (bad news). Moreover, while the cost of a negative information shock is immediate, the benefits, in terms of improved decisions, are frequently delayed.

Such a desire to avoid information has two key implications for public policy. First, information-based interventions may impose an additional cost that has not yet received attention in the literature. Consider a "nudge" in which residential energy consumers are informed of their neighbors' consumption (Allcott and Rogers 2014). Traditional cost-benefit analysis would weigh the resulting reduction in energy use against the cost of mailing customized utility bills. In such a comparison, the intervention turns out to be extremely cost effective (Benartzi et al. 2017). However, learning that one's own energy consumption considerably exceeds that of others may also impose a hedonic cost on the recipient. Indeed, this may be why the intervention works in the first place. Such a cost would ultimately need to be accounted for as well. The Information Preferences Scale may help detect for whom these hedonic costs are greatest. Second, policies are generally applied indiscriminately. For example, all customers of a utility provider receive an identically designed bill – and all employees covered by a particular health plan receive the same brochures. However, informational interventions are likely to be more effective for those who are receptive to the information, rather than motivated to avoid it. The scale could find applications in targeting individuals for whom interventions may cause the greatest change in behavior.

Chapter 2 extends the cost of information beyond hedonic implications for individuals, and proposes that learning about complementary policies can crowd-out the desire to implement policies expected to be most effective. When facing major policy challenges, like ensuring a stable income in retirement or providing access to health care, we have accepted sweeping and costly policy changes. The social security program, for example, forgoes a restrained approach in favor of a mandate to save. Although such a policy may not be optimal for consumption smoothing, it has substantially reduced poverty in old age (Engelhardt and Gruber 2004). The Affordable Care Act similarly removed choice about whether to obtain health insurance coverage and imposed a mandate to do so. Countries around the world are debating similarly sweeping legislation to combat the threat of global climate change, although with limited success to date.

One reason to implement such paternalistic policies, even in countries that tend to put more weight on individual liberties, is that they appear to be the only solutions that effectively tackle the problem. However, since the release of the book "Nudge," (Thaler and Sunstein 2008) softer policies have shown to be effective and virtually costless to implement. Changing a retirement savings plan from an "opt-in" model, in which employees have to make a deliberate choice to join, to "opt-out," in which they are enrolled and contributing by default, has lead to considerable increases in how much people save (Madrian and Shea 2001). The "Libertarian Paternalism" approach to policy (Camerer et al. 2003a; Thaler and Sunstein 2003) has gained support around the world, with nudge units operating out of governments, international organizations, and, increasingly, the private sector ("Policymakers Around the World Are Embracing Behavioural Science" 2017).

Although nudges are extremely cost-effective, there are important situations in which their absolute effect size remains small. Nudges that promote energy conservation, for example, lead residential energy users to cut back by just over a percentage point (Allcott 2015), which is unlikely to make a sizable impact on carbon emissions, since only about 10% of such emissions are caused by residential energy use in the first place. Of course any reduction is better than nothing, especially when it indeed can be achieved at low cost. However, as I show in Chapter 2, there is a largely unrecognized and potentially pernicious cost of introducing such an environmental nudge: people may come to see it as an alternative to more substantive policy, rather than the complementary role it would ideally serve. Participants across 6 experiments, including a sample of policymakers, were less likely to support a carbon tax when they also had the option of implementing a green energy nudge. That is, the (small) gains of nudges may come at the expense of decreased support for more painful, but ultimately more impactful policies.

Finally, there is one domain in which policymakers *have* considered the cost of information: they have worried that transparency about the use of nudges could undermine their effectiveness (Bovens 2009; House of Lords 2011). If defaults, for example, work only when people are unaware that an attempt is being made to influence them, then transparency may render the nudge ineffective. Worse, learning that one has been influenced could potentially lead to backlash, making the nudged option even less likely to be chosen.

Chapter 3 reports evidence that the effectiveness of a default nudge is indeed not diminished when the intervention is transparently disclosed. Respondents in an experimental survey were asked to complete a hypothetical advance directive, in which they expressed their preferences for receiving (or declining) life-extending measures that may impose considerable discomfort, such as receiving nutrition via a feeding tube when they are unable to eat on their own. They were randomly assigned to a default of receiving or declining these interventions. I observe a default effect even when participants were aware that their default had been randomly assigned and that others had been randomly assigned to the other option, removing any recommendation signal that the default may convey.

Together, the three essays suggest that deciding whether to provide information to people in the context of public policy is not straight-forward. When a standard economic account suggests people should be informed, they may desire to remain ignorant (Chapter 1). When information about a new policy falsely implies the possibility of a quick-fix, ignorance may lead to better policy outcomes (Chapter 2). And when we might worry that disclosure undermines a policy's effectiveness, we may do so for no good reason (Chapter 3). Experimental methods, as employed in the following chapters, can advance a psychologically grounded understanding of information and set a path to more effective policymaking.

1 Measuring Information Preferences

1.1 Introduction

We live in an unprecedented age of information. Advances in genetic testing can reveal conditions decades before symptoms emerge, calories in our meals are prominently displayed on menus, and social media “likes” tell us how receptive others are to the thoughts we share. Much of this information is available at little or no (financial) cost and can be consequential for the decisions individuals make. Conventional economic models, dating back to George Stigler’s seminal paper on information as a scarce resource (Stigler 1961), suggest that decision-makers would be eager to obtain such news and make full use of it. At worst, information that turns out not to be useful can simply be ignored.

Contrary to this perspective, a substantial body of experimental and field evidence finds that people are often unwilling to learn information that could be painful. Oster, Shoulson, and Dorsey (2013), for example, find that only 7% of individuals at high risk for Huntington’s disease elect to find out whether they have the condition, despite the availability of a genetic test that is generally paid for by health insurance plans. Ganguly and Tasoff (2016) find that participants in a laboratory experiment are willing to forgo part of their earnings in order to not learn the outcome of a test for a sexually transmitted disease and that such avoidance is greater when the disease is more severe. Information avoidance is not limited to health decisions. Sicherman et al. (2016) show that investors are less likely to log on to their stock portfolios on days when the market is down, when they might expect to observe losses in their own investments. Similarly, potentially unpleasant information about one’s personal characteristics may be avoided. Eil and Rao (2011) find, in a laboratory experiment, that many participants who expect to be rated as relatively less attractive or intelligent compared to other participants are willing to pay to avoid learning their true rank. Across contexts, people appear to deliberately and actively avoid information, even when it could be instrumental and lead to better decisions.

What could explain such avoidance? Recent models of belief-based utility propose that people derive value not merely from their consumption, but also from their beliefs about themselves and the world, as well as their expectations about the future (Falk and Zimmermann 2014; Koszegi and Rabin 2006; Loewenstein 2006). That is, information itself can have hedonic costs and benefits that have to be traded off against the decision utility of the information. When decision-makers fear that the information could be unfavorable, they may decide to not obtain it in an effort to protect the value they derive from their (potentially false) belief, even as that may undermine the quality of subsequent decisions. For example, learning one's level of attractiveness does not merely provide value because it informs other decisions (e.g., whom to invite out on a date), but learning that one is (un)attractive may provide (dis)utility regardless of whether the information changes one's decisions. The mechanism of anticipated regret (Zeelenberg 1999), whereby we imagine a better outcome had an alternative been chosen, may cause people to avoid information expecting that they will regret knowing the truth.

Failure to obtain information can have implications for society at large. Communicable diseases such as HIV may fail to get diagnosed and proliferate as a result (Caplin and Eliaz 2003; Sullivan et al. 2004). Voters may not consider information that challenges their ideological views, potentially causing insufficient and biased updating that may contribute to political polarization (Druckman, Peterson, and Slothuus 2013). In the case of climate change, active avoidance of scientific consensus may contribute to policymakers' failure to take actions to deal with the problem (Ho, Budescu, and Por 2017; Marshall 2014). In an organizational context, managers at firms may (deliberately) fail to learn about ethical transgressions of their employees (Bazerman and Sezer 2016), with costly consequences for society as well as, often, the firm itself.

Of course, not all individuals avoid potentially unpleasant information in all situations. Some people routinely get tested for sexually transmitted diseases or expose themselves to political views contrary to their own. This suggests that information preferences may be an important source of individual differences, similar to time and risk preferences. Unlike for those two important characteristics, however, there is no commonly used measure to assess preferences for information. Indeed, despite the many serious consequences that avoiding information may have for society or the individual, we know little about who these avoiders are, and hence cannot identify them in empirical research or develop potential interventions that target them. Although economics and psychology both offer potential explanations for information avoidance (Golman, Hagmann, and Loewenstein 2017; Sweeny et al. 2010), to date there has been no empirical work clarifying information preferences as a psychological construct. Existing studies primarily test one-time, context-specific decisions and, with one exception that we discuss below (Howell and Shepperd 2016), there is no direct method of eliciting such individual preferences across a variety of situations. This leaves unanswered questions about the generality of information aoidance across domains, its prevalence, and its consequences.

In this paper, we develop and validate a scale measuring information preferences. Our scale asks respondents to imagine themselves in a series of hypothetical scenarios in which they can choose to obtain (or not obtain) information. The scenarios cover three domains that span many high-stakes decisions, and for which there exists empirical evidence of avoidance: health, e.g. obtaining an estimate of one's life expectancy; finance, e.g. learning about the performance of alternative investments that one could have pursued; and personal characteristics, e.g., how attractive others believe one to be. We rely on scenarios to make salient the potential hedonic cost of obtaining the information. This is a deviation from scales measuring other constructs that rely on abstract questions and we show that our scenarios better predict consequential information acquisition decisions than do abstract questions.

We first outline the development of the Information Preferences Scale (IPS) building on insights from four pilot studies. Then, in Study 1, we identify the latent factors underlying information preferences and show prevalence of information avoidance across a variety of scenarios and domains. We also compare the discriminant and convergent validity of information preferences with established measures of related theoretical constructs. We predict that information preferences will differ across domains. For example, individuals afraid of learning potentially negative health news may not be averse to learning about their attractiveness. In Study 2, we confirm the latent factor structure of information preferences on a new sample using a confirmatory structural equation model, and we verify test-retest reliability of the scale via two scale administrations that were four weeks apart. Using the theoretical constructs most related to information preferences from the prior study, we refine the conceptual (dis)concordance of information preferences by further comparing the IPS with additional established measures.

Study 3 further tests the extent to which the scale can predict real-world decisions to acquire information in all three domains, and provides an additional test of convergent validity by comparing our measurement to an alternative scale designed to measure information avoidance using abstract questions, rather than specific scenarios (Howell and Shepperd 2016). Study 4 tests the scale's ability to predict information acquisition decision in a domain different from any of the three included in the scale itself. We show that our scale predicts not just intentions, but actual behaviors related to information acquisition and conclude with recommendations for applications of the scale.

1.2 Scale Development

Given the difficulty of capturing the diverse and almost infinite situations in which one seeks, or more interestingly, avoids information, our scale development process focuses on three domains in which information avoidance has been empirically demonstrated and which plausibly provide information that people may be motivated to avoid: health, consumer finance, and personal characteristics. These are topics for which information of uncertain valence may induce anxiety and discomfort but for which attaining more accurate beliefs can yield considerable benefits. Early health interventions can extend life expectancy, learning about financial mistakes can improve future financial well-being, and accurate information about how one is perceived by others can improve self-presentation and social interactions. The three domains allow us to cover a broad range of information acquisition decisions and to explore whether avoidance in one domain (e.g. health) is also predictive of avoidance in another domain (e.g. finance).

Ajzen and Fishbein (1977) point out that attitudes and behaviors often do not correspond unless the attitude is related to the behavior. To that end, rather than elicit attitudinal dispositions, we design each item to contain a specific hypothetical scenario in which a decision-maker has an option to acquire potentially useful information, though perhaps at a risk of a negative surprise (e.g. learning one has made a mistake in the past). Learning an outcome can increase the quality of later decisions, but at a possible emotional expense.

The scenarios are written to represent situations that people may typically encounter and may already have experience with, e.g. whether to look at the performance of an investment opportunity they did not pursue. This increases content validity, or the extent to which the scale is representative of a general population's experiences. To make the framing more natural and to minimize asking leading questions that might exaggerate information avoidance, all items ask about the desire to obtain the information (rather than avoid it).

In our first pilot study, we categorized people according to a four-fold classification of information preferences by giving participants in each scenario the choice of whether to either completely avoid an item of information, avoid the information only if a) they expect a negative outcome (e.g., to not look at credit score if they suspect it is low), b) avoid the information only if they expect a positive outcome (e.g., that they are viewed as more attractive than they thought), or c) seek information regardless of their expectations. Some items also tapped into the temporal aspect of avoidance: the choice to delay, but not entirely avoid consumption of information, e.g., by setting aside an envelope with a bill to be opened at a later date. A general information preferences question described the tendency for people to avoid information when it could be painful or seek it even when it may be painful, and asked participants to rate themselves along this continuum.

Respondents were generally less receptive to information when they expected negative outcomes, confirming similar experimental evidence (Oster, Shoulson, and Dorsey 2013; Ganguly and Tasoff 2016; Eil and Rao 2011). We retained pilot items if they exhibited a biserial correlation of greater than 0.25 for at least one part of the question with both the general information preferences question and the total sum-score. Items examining delay in information seeking were not predictive of either criterion and were excluded, along with three items that describe situations less commonly encountered outside of the United States that could have restricted international usage of the scale.

In the second iteration, participants evaluated separately whether they would obtain or avoid the information in two circumstances: when the expected outcome was positive and when it was negative. Again, participants typically reported more avoidance with a negative expected outcome. As a higher proportion of participants avoided information when it was expected to be negative, we rewrote and tested a new set of general items which measured the inclination to remain ignorant in a situation even when others may know bad news about the individual.

The penultimate pilot study tested a revised set of items such that no outcome (positive or negative) was explicitly stated, but the possibility of either outcome was implicit. The new items, in line with previously generated successful items, sought to capture universal experiences and situations (e.g., whether to check if your recommendation to a friend was well-received). The four-fold classification was initially distilled into a binary decision: simply the decision to acquire or avoid information. However, a final pilot study that used a four-point ordinal response scale yielded higher internal consistency (measured by Cronbach's α) than when participants were presented only with a dichotomous choice; hence, the final resulting scale incorporates the ordinal responses.

In all the scenario-based questions, the information is depicted in a way so that (1) information is of uncertain valence, i.e. it could be favorable or unfavorable, and (2) the potential discomfort is experienced when the information is obtained while the potential benefits of obtaining the information are in the future. Prevalence of information avoidance is defined by proportion of respondents who definitely or probably did not want to know a piece of information. The final scale contains 13 items (5 personal characteristic items, 3 health items, 3 finance items, and 2 general items; Table 1).

Table 1: Scale Items and Proportion of Avoiders.

Domain	Scale Item	Avoidance
Health	As part of a semi-annual medical checkup, your doctor asks you a series of questions. The answers to these questions can be used to estimate your life expectancy (the age you are predicted to live to). Do you want to know how long you can expect to live?	43.68%
	You provide some genetic material to a testing service to learn more about your ancestors. You are then told that the same test can, at no additional cost, tell you whether you have an elevated risk of developing Alzheimer's. Do you want to know whether you have a high risk of developing Alzheimer's?	21.32%
	At your annual checkup, you are given the option to see the results of a diagnostic test which can identify, among other things, the extent to which your body has suffered long-term effects from stress. Do you want to know how much lasting damage your body has suffered from stress?	25.52%
Finance	Ten years ago, you had the opportunity to invest in two retirement funds: Fund A and Fund B. For the past 10 years, you have invested all your retirement savings in Fund A. Do you want to know the balance you would have, if you had invested in Fund B instead?	51.05%
	You decide to go to the theater for your birthday and give your close friend (or partner) your credit card so they can purchase tickets for the two of you, which they do. You aren't sure, but suspect that the tickets may have been expensive. Do you want to know how much the tickets cost?	20.52%
	You bought an electronic appliance at a store at what seemed like a reasonable, though not particularly low, price. A month has passed, and the item is no longer returnable. You see the same appliance displayed in another store with a sign announcing 'SALE.' Do you want to know the price you could have bought it for?	37.37%
Personal Characteristics	You gave a close friend one of your favorite books for her birthday. Visiting her apartment a couple of months later, you notice the book on her shelf. She never said anything about it; do you want to know if she liked the book?	23.68%
	Someone has described you as quirky, which could be interpreted in a positive or negative sense. Do you want to know which interpretation he intended?	31.31%
	You gave a toast at your best friend's wedding. Your best friend says you did a good job, but you aren't sure if he or she meant it. Later, you overhear people discussing the toasts. Do you want to know what people really thought of your toast?	39.21%
	As part of a fund-raising event, you agree to post a picture of yourself and have people guess your age (the closer they get, the more they win). At the end of the event, you have the option to see people's guesses. Do you want to learn how old people guessed that you are?	24.21%
	You have just participated in a psychological study in which all the participants rate one-anothers' attractiveness. The experimenter gives you an option to see the results for how people rated you. Do you want to know how attractive other people think you are?	39.21%
General	Some people seek out information even when it might be painful. Others avoid getting information that they suspect might be painful, even if it could be useful. How would you describe yourself?	28.69%
	If people know bad things about my life that I don't know, I would prefer not to be told	33.69%

1.3 Study 1

Study 1 explores the latent factor structure underlying the measure of information preferences devised in earlier pilot studies. Additionally, we examine the relationship between the Information Preferences Scale (henceforth IPS) and other conceptually-related measures. Experiments frequently measure participants' time and risk preferences and we expect those measures to correlate with the desire for information. Because the psychological cost (e.g., anxiety, disappointment) occurs immediately and the benefits (e.g., realization of better financial decisions) occur in the future, those who discount the future more are expected to also be less likely to desire information. We predict that those with high discount rates overweigh the anticipated pain that comes with immediately consuming the information, and underweigh the positive utility that such knowledge may bring in the future. Similarly, because the valence of the information is uncertain, individuals face the prospect of either learning favorable news or unfavorable news. That is, the decision to acquire information is similar to that of a risky lottery, with some probability of a gain and some chance of incurring a loss. We hypothesize that individuals who are more tolerant of risk are then also more willing to obtain information.

1.3.1 Method

1.3.1.1 Subjects

We recruited 400 participants (52.89% male with a mean age = 34.92, *SD* = 9.99) via Amazon Mechanical Turk. Of those, 18 participants failed the attention check and 2 did not complete the full survey. We analyze data from the remaining 380 participants.

1.3.1.2 Procedure

Participants completed the 13-item scale generated from the pilot studies. To ensure consistency in response format across items, the response scale for all items was on a 1-4 Likert scale, from 1 =

"Definitely don't want to know" to 4 = "Definitely want to know." For each item, we standardized the response by subtracting from it the mean and dividing it by the standard deviation. To assess the relationship between information preferences and other established constructs hypothesized to be related to our measurement, participants also completed measures for risk aversion (Gneezy and Potters 1997), time preferences (Kirby, Petry, and Bickel 1999), openness to opposing views (Minson, Tinsley, and Chen 2018), preference for coherence (Antonovsky 1993), need for cognition (Cacioppo, Petty, and Kao 1984), preference for consistency (Cialdini, Trost, and Newsom 1995), as well as several general personality traits (BFI; John and Srivastava 1999). We presented the measures and scales (including our own) in random ordering to prevent order effects. Items in each measure were appropriately reverse-coded and, with the exception of the time discounting task, averaged to produce a mean score. The time discounting measure was calculated by identifying the point of indifference between two valuations (see Kirby, Petry, and Bickel 1999).

1.3.2 Results

Across all scenarios we observe a considerable degree of avoidance, suggesting that information avoidance is highly prevalent. On average across all participants and items, 32.46% of responses indicated a definite or probable preference for not obtaining the information. The fraction of avoidant responses across items ranged from 20.52% to 51.05% (Table 1). To produce an information preferences score, items were averaged (M_{score} = 0.67, SD_{score} = 0.16). The distribution of scores for all studies are shown in Figure 1.

To determine whether demographic variables influence information preferences, we regress gender, education, political affiliation, income, and age on the scale scores. No coefficient was significant at the α = 0.05 level; nor is the resulting model, $F(25, 354) = 0.73$, p = 0.82. This suggests that information preferences do not differ across any broadly defined demographic group.

1.3.2.1 Exploratory Factor Model

Figure 1: IPS score distribution density plots with median (Studies 1-4).

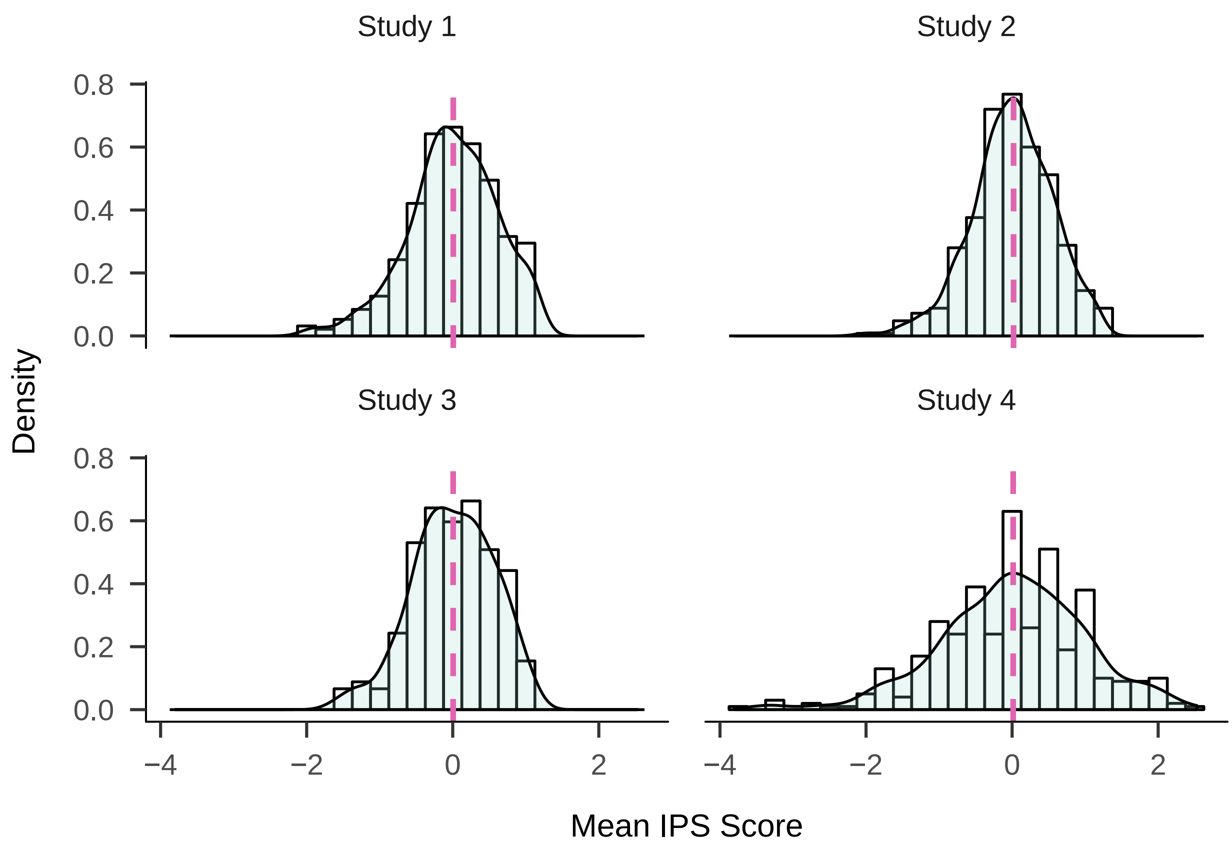


Table 2: Standardized factor loadings for the EFA (Study 1) and CFA (Study 2).

	Item	EFA Loadings	CFA Loadings
Health	H1	0.66	0.71 ***
	H2	0.80	0.72 ***
	H3	0.75	0.69 ***
Finance	F1	0.37	0.64 ***
	F2	0.29	0.27 ***
	F3	0.49	0.70 ***
Personal	I1	0.96	0.40 ***
	I2	0.54	0.59 ***
	I3	0.59	0.56 ***
	I4	0.73	0.66 ***
	I5	0.75	0.68 ***
General	G1	0.81	0.81 ***
	G2	0.81	0.67 ***

[a] . $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
[b] EFA = Exploratory Factor Analysis.
[c] CFA = Confirmatory Factor Analysis.

To examine the latent factor structure of information preference, we perform exploratory factor modeling on the scale in a two-step procedure. First, to determine the number of latent factors, we apply Kaiser's rule (1960), which retained 4 latent factors from the scale's 13 items. We hypothesize that information preferences consist of three domain factors as well as a general factor. Then, we fit an exploratory factor analysis on the 11 domain items using an oblimin factor rotation, which accounts for correlations across latent factors. A three-factor model provides the best model fit. As predicted, the items have high loadings on their intended domains, e.g., all health items cluster together to form an individual factor. This implies that the propensity to obtain information is conditional on the particular topic. The two general items exhibited moderate correlations with items from all the three domains. To incorporate the additional two general items, we fit an exploratory structural equation model (Asparouhov and Muthén 2009) on all 13 items into a general factor whilst simultaneously accommodating the three-factor structure uncovered in the domain items. The exploratory factor loadings are presented in Table 2. We verify this model in a confirmatory analysis in Study 2.

1.3.2.2 Divergent Validity

To examine the divergent validity of the 13-item scale, we compare the correlations between established measures, the IPS, and the domain-specific items (Table 3).

Table 3: Divergent validity correlations (Studies 1 and 2).

	Comparison Scale	Health	Consumer Finance	Personal	Sum Score Total
Study 1	Need for Consistency	-0.06	0.03	-0.03	-0.07
	Need for Closure	-0.07	0.04	0.04	-0.12 *
	Receptiveness to Opposing Views	0.13 **	-0.02	0.09 .	0.23 ***
	Need for Cognition	0.12 *	0.09 .	0.15 ***	0.21 ***
	General Risk	0.05	0.07	0.13 **	0.12 *
	Time Discounting	-0.07	-0.08	-0.12 *	-0.16 ***
	BFI: Extraversion	0.00	0.11 *	0.13 **	0.11 *
	BFI: Agreeableness	0.02	-0.04	0.07	0.04
	BFI: Conscientiousness	0.03	0.13 **	0.04	0.14 **
	BFI: Neuroticism	-0.03	-0.08 .	-0.03	-0.17 ***
	BFI: Openness	0.18 ***	0.10 *	0.18 ***	0.22 ***
Study2	Curiosity	0.13 ***	0.03	0.13 ***	0.22 ***
	Self-Efficacy	0.14 ***	0.08 *	0.18 ***	0.21 ***
	Learning Styles	0.23 ***	0.11 **	0.25 ***	0.31 ***

1.3.2.2.1 Preference for cognitive activities

To examine information preferences are related to a propensity for satisfying other types of knowledge gaps, we examined the correlation between our scale and the Need for Cognition scale (NFC; Cacioppo, Petty, and Kao 1984). The correlation between information preferences and the NFC was positive, $r(379) = 0.21$, $p < 0$, indicating that those with a high need for cognition also have a tendency to desire information (Table 3). The Receptiveness to Opposing Views scale (Minson, Tinsley, and Chen 2018) assesses the tendency to listen to opinions that are contrary to one's own, closely-held beliefs. As one might expect, participants who preferred information in general were also more likely to be receptive to hearing viewpoints that differed from their own, $r(379) = 0.23$, $p < 0$.

Need for Closure (Webster and Kruglanski 1994) measures a preference for order, structure, and predictability, over ambiguity (Kruglanski 2013). We hypothesize that those exhibiting a great need for closure would be more willing to disregard evidence that either does not correspond with already formulated opinions or induces re-evaluation. We observe a low but significant negative correlation between Need for Closure and the IPS, $r(379) = -0.12$, $p < 0.05$, suggesting that those who prefer order and structure are more likely to avoid psychologically discomfiting information. The Preference for Consistency-Brief Scale (PfC-B; Cialdini, Trost, and Newsom 1995) was not correlated with the IPS, $r(379) = -0.07$, *ns*, perhaps because the PfC-B scale measures both an individual preference for consistency and also a self-reported perception of how others see one in this regard, whereas the IPS measures only the individual trait.

1.3.2.2.2 Risk, Time, and Information Preferences

Receiving information of uncertain valence can be risky: it could turn out favorably, with news better than expected. However, it could also be worse than expected, leading to a loss in belief-utility. Consequently, we hypothesized that people who are more open to taking risks may also be more willing to obtain information that could be either positive or negative. Indeed, consistent with this account, we find that risk tolerance is positively correlated with the desire to obtain information $r(379) = 0.12$, $p < 0.05$.

Similarly, as obtaining potentially painful information (e.g., watching a video of oneself giving a talk) often involves an immediate cost in exchange for a delayed gain (e.g., improved teaching in the future), we also anticipated that information avoidance would be associated with high rates of time discounting. Analogously and in line with prior work, temporal discounting might also influence people's decision to obtain information immediately or delay consumption to the future (Falk and Zimmermann 2014). As predicted, we observe a negative relationship between the desire to obtain information and the discount rate (Kirby, Petry, and Bickel 1999), $r(379) = -0.16$, $p < 0$.

1.3.2.2.3 General Personality Traits

We look at the relationship between information preferences and the Big Five Personality Inventory (BFI; John and Srivastava 1999). The desire for information was uncorrelated with agreeableness ($r(379) = 0.04$, *ns*), but positively correlated with extraversion ($r(379) = 0.11$, $p < 0.05$, conscientiousness ($r(379) = 0.14$, $p < 0.05$), and openness to new experiences ($r(379) = 0.22$, $p < 0$). information preferences are negatively correlated with neuroticism ($r(379) = -0.17$, $p < 0$). Extraversion, characterized by high sociability and expressiveness, may induce those exhibiting high levels of this trait to also seek information more. Exhibiting high conscientiousness, i.e., a tendency towards perseverance, may be the counterpoint against information-delaying inclinations.

Conversely, neuroticism, the tendency to more readily experience unpleasant emotions, may increase the hedonic cost of obtaining unfavorable information and hence make one less likely to take a chance in obtaining it. People who score high on the openness to new experiences factor, which relates to a tendency towards intellectual pursuits, also score high on curiosity (John and Srivastava 1999), and may incur a cost from *not* having information that they know is available, irrespective of its valence.

1.3.3 Discussion

Study 1 examines the factor structure of the IPS and its relationship to a broad range of other measured constructs. In a purely exploratory model, the domain items all load onto their respective latent factors (e.g., health items all mapped onto the same factor), providing a clear multi-dimensional factorial structure of information preference. This result implies that information preferences are sensitive to the context in which the information is embedded, providing support for our second hypothesis that information preferences are sensitive to domain. Yet, we also sought to capture a more general and contextless aspect of information preferences with our two general items, and the exploratory factor model fitted suggests the latent factor structure of information preferences can accommodate both individual personality differences and context-dependent dimensions (Mischel and Shoda 1995).

To the best of our knowledge, this is the first empirical evidence comparing within-subjects' differential propensities towards obtaining information; previous studies (e.g., Sullivan et al. 2004) have focused instead on specific, one-time situations where usually only a single decision is involved. We see sizable proportions of avoidance across the wide variety of situations depicted in the IPS. To the extent that self-reporting behavior introduces bias (e.g. because participants want to project a favorable view of themselves), and to the extent that information-seeking is viewed as normative, we are if anything underestimating the extent of avoidance.

The discriminant validity of information preferences' psychological uniqueness is affirmed by its lack of correspondence with potentially related constructs such as measures of need for consistency, closure, cognition, risk attitudes, receptiveness to opposing views, time discounting, and general personality traits. The scale appears to measure a distinct construct, with none of the correlations between the scale and potentially related measurements exceeding an absolute value of 0.3. Additionally, information preferences are not predicted by standard demographic characteristics such as gender, income, age, or political affiliation. In concert, these findings suggest the factors underlying the latent construct of information preferences are unique and cannot be explained solely by existing measurements. To further confirm and replicate these results, in Study 2 we administer the IPS to another sample at two time points, allowing us to assess test-retest reliability as well as providing additional, empirically motivated tests of convergent and discriminant validity.

1.4 Study 2: Test-retest reliability

We confirm the proposed exploratory factor model from Study 1 with a new and larger sample. By eliciting the scale from the same respondents at two points in time, we were able to assess test-retest reliability. We further test, beyond the measures in Study 1, the discriminant validity of the information preferences scale by comparing it with additional constructs that have potential theoretical overlap. This allows a further clarification of the correspondence between information preferences and other established personality traits. We selected measures that bore the most theoretical similarity to the constructs most highly correlated with information preferences in Study 1: curiosity, self-efficacy, and different learning styles. For example, a moderate correlation between information preferences and openness to new experiences in Study 1 suggests that information preferences may also be linked to curiosity, which is typified by search for information that may not be particularly useful (Loewenstein 1994). Recently, receptiveness to oppositional political views has also been linked to curiosity (Kahan et al. 2017), further lending support to a potential relationship between preferences for information and curiosity.

1.4.1 Method

1.4.1.1 Subjects

We recruited 601 participants (52.8 % male with a M_{age} = 36.71, SD_{age} = 12.01) on Amazon Mechanical Turk to complete the scale at two time points about four weeks apart. To avoid biasing our results, we report results for the 500 participants who completed both stages in our analysis. Those who failed to respond to the follow-up survey do not differ on any demographic measure from those who did complete the follow-up.

1.4.1.2 Procedure

To examine the stability of the psychological trait over time, participants completed the assessment twice, with a four-week lag between the two administrations. For the first administration of the IPS only, we included additional psychological measures: the Curiosity and Exploration Inventory (CEI-II; Kashdan et al. 2009) and the General Self-Efficacy Scale (GSE; Schwarzer and Jerusalem 1995).

1.4.2 Results

We observe high internal consistency in both the first (measured by average inter-item correlations; Cronbach's $\alpha = 0.8$) and second ($\alpha = 0.83$) administration of the IPS. Test-retest reliability, measured by the correlation of respondents' average scores across both time points, was $r = 0.64$, indicating that the IPS reliably measures the construct over time.

1.4.2.1 Confirmatory factor analysis

We fit a confirmatory structural equation model on the responses for the first administration of the scale. Due to the ordinal (non-continuous) response type, the IPS is not normally distributed, $M_{\text{skewness}} = 1188.56$, $p < 0$, $M_{\text{kurtosis}} = 17.99$, $p < 0$ (Mardia 1970). In such a situation, Floran and Curran (2004) recommend using the diagonal weighted least squares estimation procedure to estimate the confirmatory latent model.

Confirming the exploratory factor model in Study 1, the resulting latent factor structure (Figure 2) contains four correlated factors: the three domains and a general information preferences factor. The general factor loads onto the latent domains as well as the two general items (Table 2). The root mean square error approximation (RMSEA), a model fit index (Steiger and Lind 1980), is $\epsilon = 0.03$, 90% confidence interval, [0.02, 0.04], and falls within guidelines of good model fit (< 0.08; (Hooper, Coughlan, and Mullen 2008). This is corroborated by other fit statistics, Tucker-Lewis Index = 0.98 (Tucker and Lewis 1973), and Comparative Fit Index = 0.99, above recommended cutoffs of 0.90 and 0.95, respectively (Hu and Bentler 1999). The latent factor correlations are shown in Table 4.

Figure 2: SEM Plot (Study 2).

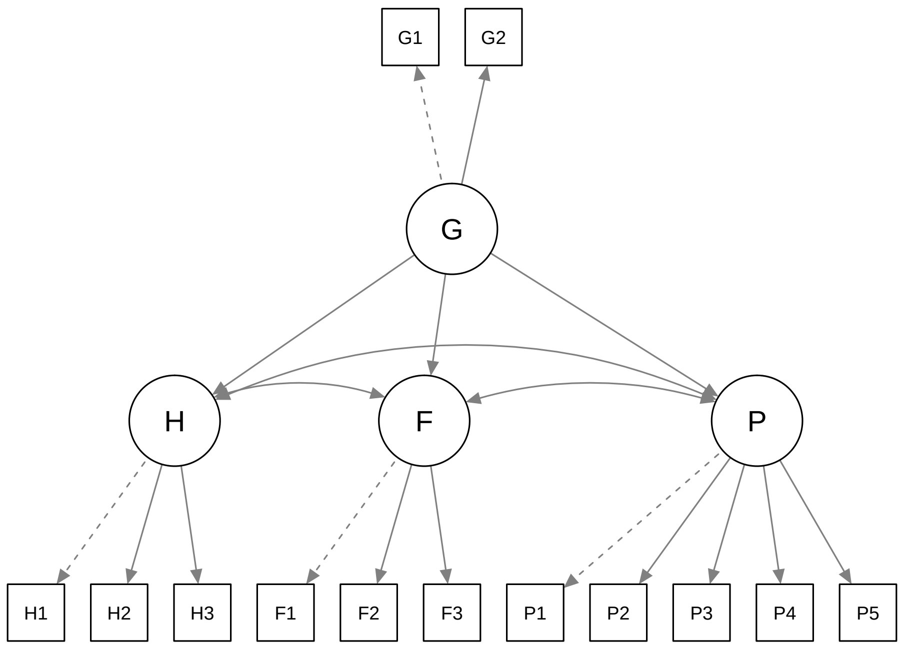


Table 4: Latent factor correlations (Study 2).

	Health	Finance	Personal	General
Health	1			
Finance	0.39	1		
Personal	0.54	0.48	1	
General	0.76	0.44	0.53	1

1.4.2.2 Curiosity, Self-Efficacy, and Learning Style

We further assess the theoretical correspondence between information preferences and other personality traits. As predicted, curiosity was positively related to the desire for more information, ($r(585) = 0.22$, $p < 0$). Conscientiousness, which relates to being motivated and persevering, exhibited a positive relationship in Study 1 with self-efficacy, or the extent to which individuals believe themselves to be capable of performance on a task, both academically (Caprara et al. 2011) and in the workforce (Martocchio and Judge 1997; Lee and Klein 2002). Such efficacious attitudes are linked to goal-oriented behaviors and motivation (Bandura 1986), and Schunk (1990) has suggested that those with low reported self-efficacy may avoid learning to prevent confirming personal suspicions of inadequacy. We hypothesized that self-efficacy may influence the perceptions of information usefulness; that is, those exhibiting high self-efficacy would feel more confident in their ability to make better decisions in the face of potentially negative information and thus be more likely to obtain such information. We see a positive relationship between information-seeking preferences and the GSE, $r(585) = 0.21$, $p < 0$.

1.4.3 Discussion

Study 2 demonstrates the psychometric stability of the scale over time. In addition, using a latent model approach, our confirmatory factor model provides further evidence that IPS reliably and validly measures both domain-specific preferences for information as well as information preferences as a general psychological trait. Moreover, we find that individuals have different information-seeking preferences for health, finances, and personal characteristics. We further clarify the unique construct of information preferences as compared to other psychological constructs most closely aligned with those possessing highest convergent validity in Study 1. The correlations, while statistically significant, remain moderate (Study 1 range: [-0.17, 0.23]; Study 2 range: [0.21, 0.31], further lending evidence that the desire to seek or avoid information can be reliably measured by the IPS, and that information preferences are not simply an amalgamation of other existing constructs.

1.5 Study 3: Predicting Information Choices

Studies 1 and 2 show that the scale is reliable, contains a stable factorial structure, and possesses adequate discriminant and internal validity. Having illustrated the psychometric robustness of the scale, we next explore the potential external validity of the scale. Here, we provide a systematic test of the IPS and all its subscales on domain-related information seeking and avoiding behaviors. We test, on a more diverse sample, whether our validated scale can predict behaviors associated with information preferences across all domains represented in the IPS. In line with evidence from the first two studies suggesting that construct of information preferences are domain-specific, we hypothesize that the scores from the domain-specific items can predict a related decision to obtain information. Moreover, as the IPS was designed to be predictive across a variety of contexts, we also hypothesized that scores from the complete IPS would predict information acquisition decisions across domains. Additionally, we compare our scale to an alternative assessment measuring information avoidance using abstract questions (Howell and Shepperd 2016), providing an additional test of convergent validity and a benchmark for our scenario-based approach.

1.5.1 Methods

1.5.1.1 Subjects

We recruit readers of a science message board **Reddit r/science** via a discussion about information avoidance and listeners of a behavioral science podcast *You're Not So Smart* in an episode on the same topic. We thus reached a sample that was both geographically diverse (40.33% from outside the US) and highly educated (32.04% with graduate degree). In total, 181 participants (61.33% male with a M_{age} = 35.57, SD_{age} = 11.28) completed our study.

1.5.1.2 Procedure

Participants completed the 13-item scale and were asked to make a consequential decision to obtain (or avoid) information in one of the three domains represented in the scale. If participants chose to obtain the information, they were forwarded to the relevant website upon completion of the study. Participants were randomly assigned to make this decision in the health domain (a website calculating their life expectancy), the personal characteristic domain (an algorithm that would estimate their age from a picture), or the financial domain (a website to estimate their income in retirement). We also varied whether we highlighted potentially positive information (e.g. you might live longer than you expected) or negative information (e.g., you might not live as long as you expected). Because fewer respondents completed the survey than expected, we are not sufficiently powered to detect differences in that framing and combine both in our analyses. Because the issue of retirement savings is less relevant in many countries outside of the United States, non-US participants were not assigned to this condition. Finally, participants also completed a recently published alternative measure for information preferences (HS; Howell and Shepperd 2016). The format of HS is open-form (example item: “I would rather not know ______”). These sentence stems allow the test-maker to complete the sentence as they wished. In line with the phrases used in the Howell and Shepperd (2016) study, the sentences were completed using the phrases “my health”, “my finances”, and “how attractive others find me” for the three domains. We counterbalanced the order of the questions as follows: the behavioral measure was asked either first or last and we randomized the order of the IPS and the HS scale.

1.5.2 Results

The IPS and the HS scale were moderately negatively correlated, $r(180) = -0.66$, $p < 0$, indicating an oppositional correspondence between information preferences and information avoidance, as expected. To examine the relationship between our scale and HS scores on propensity to opt for information, we conducted logistic regressions. The IPS in conjunction with the HS scale, the total IPS scale, its individual subscales, the total HS scale, and the individual HS subscales were regressed separately on the decision to seek or avoid information.

The complete IPS scale significantly predicts information seeking across all three decision tasks, $OR_{Overall} = 1.11$, 95% CI = [1.06, 1.16], $p < 0$. The HS scale is, in contrast, not significant, $OR_{Overall} = 0.99$, 95% CI = [0.98, 1], $p = 0.09$. When both scales are compared in a multiple logistic regression, the total IPS score significantly predicts information-seeking behaviors over the HS score, $OR_{IPS \mid HS} = 1.09$, 95% CI = [1.04, 1.15], $p < 0$. The HS score is non-predictive, $OR_{HS \mid IPS} = 1$, 95% CI = [0.99, 1.02], *ns*.

This suggests that the IPS exhibits ecological validity, as it is able to succinctly predict information-seeking behaviors across multiple contexts. A statistical comparison of the two scales shows clearly that the IPS can significantly explain variations in information-seeking and -avoidant behaviors beyond what the alternative scale can provide. We confirm this finding by applying dominance analysis (Azen and Budescu 2003; Azen and Traxel 2009) to the logistic regression specification containing both scales. This approach uses changes in model fit statistics (i.e., R^2 in ordinary least squares regression) to determine predictor importance. Rather than the standard variance-explained index, logistic regression relies instead on quasi-R^2 indices (Cox 1985; Estrella 1998; McFadden 1974; Nagelkerke 1991). The Nagelkerke index shows that the contribution of the IPS scale ($R^2_N = 0.05$) is larger than that of the HS scale ($R^2_N = 0$). This pattern replicates in the three other quasi-R^2 indices tested, establishing dominance of the IPS over the HS scale.

1.5.2.1 Domain-specific results

For information on one's life expectancy, having a high score on the health subscale significantly predicted the odds that participants would seek the information (rather than avoid it), $OR_{Health\ Seeking} = 1.52$, 95% CI = [1.1, 2.18], $p < 0.05$. (Equivalently, scoring low on the subscale significantly increased the odds of *avoiding* the information, $OR_{Health\ Avoidance} = 0.73$, 95% CI = [0.65, 0.82], $p < 0$.) In the personal characteristics domain, having a high score on the characteristics subscale also significantly predicted the odds of seeking information on how old the participant looked, based on an algorithm scoring a self-portrait, $OR_{Personal} = 1.31$, 95% CI = [1.09, 1.61], $p < 0.01$. In the consumer finance condition, participants scoring high on the corresponding subscale tended to also wanted to know more about their retirement savings, though this was not significant at the usual level, $OR_{Finance} = 1.32$, 95% CI = [0.98, 1.86], $p = 0.08$. Because we had a large group of international participants and fewer respondents than we had expected, we had low power only for this domain. In a similar series of logistic regressions, the HS subscales do not significantly predict any related behaviors.

For ease of interpretation, we present results of a linear probability regressions. Both the IPS and HS scales were averaged and the latter was linearly transformed so that it conformed to a 4-point scale. The reported coefficient represents the predicted percentage point increase in obtaining the information when the IPS shifts by one SD. Across all domains, with every additional one-unit SD increase in the IPS the probability of seeking information increases by 27 percentage points, $p < 0$, whereas an equivalent shift in the HS scale leads to a non-significant 11 percentage point increase in avoidance. The domain-specific linear probability models for both scales are in Table 5. The IPS significantly predicts specific domain-related information seeking or avoiding behaviors, as well as overall information seeking and avoiding across domains.

Table 5: Linear Probability Models (Study 3).

	N	IPS	HS
Health	75	0.2 **	-0.09
Finance	36	0.18 .	0.06
Personal	70	0.29 ***	-0.11
All Domains	181	0.27 ***	-0.11 .

[1]

[1]. p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001

1.5.3 Discussion

Study 3 investigates the predictive validity of the IPS with domain-specific behavioral measures. We also compare our scale with a related elicitation (Howell and Shepperd 2016). The two scales are moderately and negatively correlated, suggesting that, the scales overlap in what they measure. The IPS, however, is able to predict a variety of real-world behaviors related to information acquisition. Moreover, the subset of items pertaining to a particular domain is also able to predict the domain-specific decision to acquire information, suggesting that it may not be necessary to present all scenarios in some instances. Compared to a more abstract scale, relying on specific hypothetical scenarios, as we do in the IPS, appears to better capture the trade-off between the gains of obtaining useful information and the risk of learning something unpleasant. As Maul (2017) notes, attempting to measure noncognitive constructs by inserting phrases in open-ended sentence stems may result in statistical patterns apparently confirming validity but may lack correspondence to theoretical and behavioral outcomes. The IPS achieves both the former and the latter in its ability to use both the targeted items, as well as the entire scale, to predict behaviors related to the IPS domains. We next ask whether the IPS can predict information avoidant behavior out-of-sample, i.e., beyond domains represented in the IPS. In the next study, we expand the predictive range of the scale into the domain of politics.

1.6 Study 4

The previous study demonstrated the scale's capability to link scores to behaviors, both within domain and across the full scale. In Study 4, we test whether the IPS can predict a decision to obtain information unrelated to the domains already present in the scale. We hypothesize that, given the out-of-sample nature of the task, scores from the entire IPS would predict the decision to avoid or acquire information in the political domain.

Selective exposure to information can lead to political polarization (Druckman, Peterson, and

Slothuus 2013), so we selected politics as the out-of-sample domain. We ask participants whether they want to read a testimonial by someone affected by the Deferred Action for Childhood Arrivals (DACA) policy. DACA is an immigration policy enacted under the Obama Administration that enables undocumented immigrants who entered the country as minors to apply for a renewable two-year deferral for deportation, in addition to a permit to work. At the time the study was conducted, the Trump administration called for an end to DACA, leading to a national debate on the consequences for individuals and potential alternatives to the policy. Contrary to political polarization typically found in other immigration policies, both liberals and conservatives generally favor amnesty and pathways to citizenship for undocumented immigrants (YouGov 2017), so we suspected that both would have reasons to avoid obtaining information about it: for those on the left, it would be painful to contemplate the dismantling of a policy they supported, and those on the right may want to avoid the cognitive dissonance resulting from their party's president wanting to dismantle a policy they support (Festinger 1962). The potential existence of reasons for a broad spectrum of the population to avoid information about the policy made it fertile ground for examining information preferences. Because there are different reasons for people across the political spectrum to avoid the information, we hypothesize that the scale can predict the decision to read the article for both conservatives and liberals, but that we will nonetheless see differences by political affiliation.

1.6.1 Methods

1.6.1.1 Subjects

We recruited 400 participants (50.25 % male with a M_{age} = 35.54, SD_{age} = 10.82) on Amazon Mechanical Turk. In our sample, 26.5% of respondents identified anywhere between slightly to extremely politically conservative.

1.6.1.2 Procedure

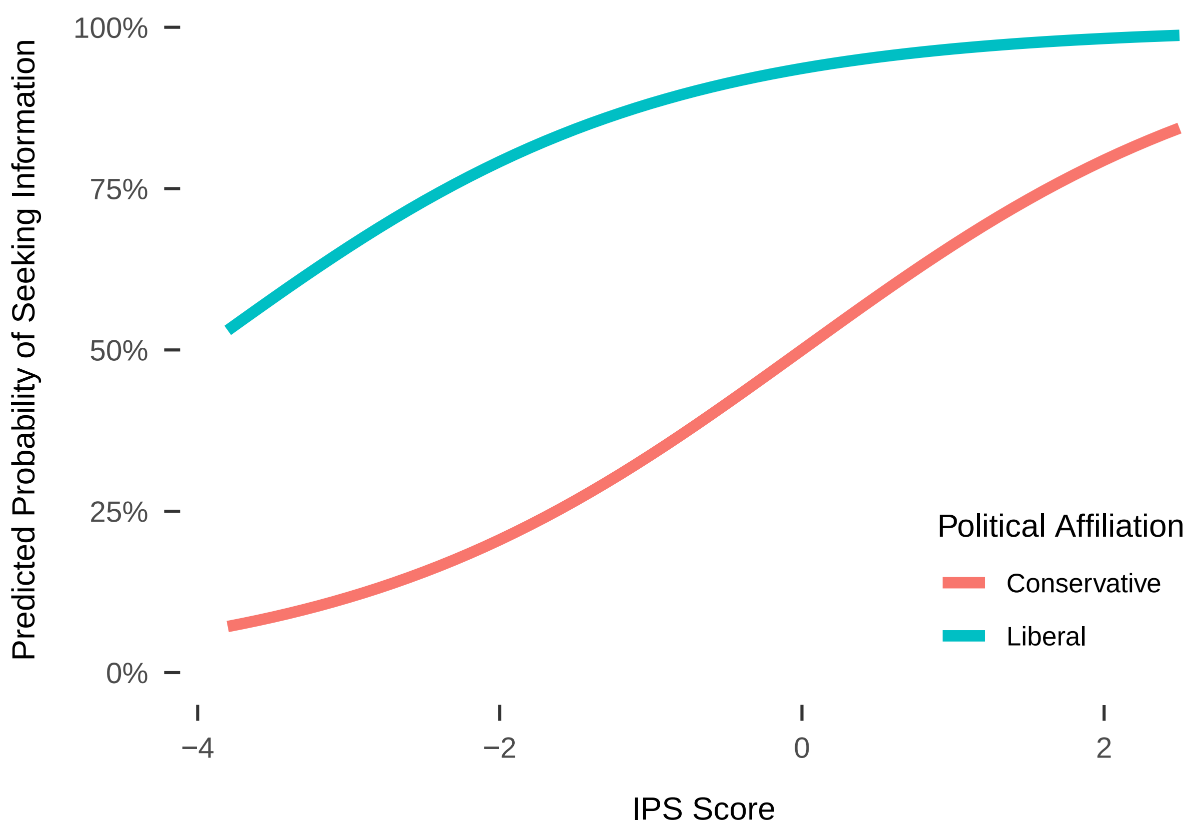
Participants read about the DACA policy and rated the extent of their support for it. We then asked them if they wanted to be forwarded to a personal testimonial from an undocumented immigrant. This article chronicled how DACA has positively impacted his life and the consequences that a policy repeal would have for him. They also completed the IPS. We counterbalanced whether participants first completed the IPS or first made the decision to be forwarded to the information. Participants then answered demographic questions. For all analyses, a median split dichotomized political affiliation.

1.6.2 Results

In our sample, only 32% of participants supported repealing DACA, and conservatives were significantly more likely to do so, $\chi^2(1, n = 400) = 59.90, p < .001$. As we found in Study 1, there was no relationship between political affiliation and IPS scores, $t(318.53) = -1.16, p = .246$.

Controlling for political affiliation and the interaction between political affiliation and the IPS scores, the IPS significantly predicts avoiding the personal testimonial, $OR_{IPS|conservative} = 0.51$, $p < 0.05$. Providing an indication of the strength of the relationship, a one-unit SD decrease in the IPS corresponds to a 33.74% decrease in the individual's likelihood of opting to obtain the information. Conservatives were 6.38% less likely than liberals to seek the information, though this effect is not significant, $OR_{conservative|IPS} = 0.07$, $p = 0.05$ (Figure 3). The interaction is non-significant ($OR = 2.2$, $p = 0.09$). The IPS on its own directionally correlated with the desire to obtain the information, but was not significant at the usual level $z = -1.72, p = .085$, underscoring the need to include other individual characteristics when predicting a behavior outside of the domains specified in the IPS. We note that including only the 11 domain items (excluding the two general information preferences items) in the a logistic regression specification including political affiliation and the IPS x political affiliation interaction also renders the effect of the scale statistically significant, with a comparable effect size; $OR_{IPS\ Domain} = 0.56$, $p < 0.05$. For the distilled scale, every one-unit SD decrease in the scale increases the likelihood of avoiding the testimonial by 36.06%.

Figure 3: Influence of political affiliation and IPS on information seeking.



Participants who elected to not read the testimonial (N = 254) gave a variety of reasons for their decision. Some participants (52.76%) expressed a belief that the link would not provide them with any instrumental information beyond what they already knew. Another 34.65% of participants believed the information would be irrelevant to them. Some participants (12.6%) also reported that they avoided the information because it would be painful or unpleasant to read.

1.6.3 Discussion

This study investigated the ability of the IPS to predict the desire to obtain or avoid information in a domain not included in the scale. We picked a narrative account related to a politically contentious topic that may be painful to read. The scale predicted the likelihood of avoiding the personal testimonial, suggesting it can generalize beyond the topics covered in the hypothetical scenarios.

Whether participants were liberal or conservative, not surprisingly, influenced the desire to read this account. However, the scale predicted information seeking behavior for both parties, suggesting that individual-level information preferences may play an important role in how political knowledge is spread (Huckfeldt and Sprague 1987). Similarly, support for the policy did not predict information seeking or avoiding, indicating prior stances and beliefs of an issue cannot explain the desire to seek out this information on their own. The IPS alone also did not significantly predict avoidance in the political domain, suggesting that combining the IPS with targeted individual characteristics, in this case political affiliation, may yield higher predictive capacity.

1.7 General Discussion

Making good decisions is often contingent on obtaining information, even when that information may be painful to think about. Substantial empirical evidence suggests that people are often ready to make worse decisions in the service of avoiding potentially painful information. We propose that this tendency to avoid potentially painful information is a trait that is separate from those measured previously, and sought to measure it using a set of items specific enough to enable respondents to imagine how they would behave if placed in the position, but sufficiently universal to capture preferences for information in a broad range of domains.

In four studies, we test the validity and reliability of the Information Preferences Scale, with a particular focus on its capacity as a behaviorally predictive tool. The IPS differs from scales that have been used to measure many other individual difference constructs in three important ways. First, it uses realistic and actionable scenarios as a foundation for defining the construct of information preference. In contrast with measures that generally rely on dispositional attributes and often lends themselves to abstracted interpretation, the IPS is oriented towards behavioral outcomes. The response type elicits the propensity that one would seek the specific information described in the item. Second, IPS items tap into a wide range of situations that are both psychologically and economically consequential in the domains of health, finance, and personal attributes. Third, the

IPS was behaviorally validated using a series of contextualized decision to acquire information across domains, and shows promise in predicting information acquisition behavior in a domain not included in the scale itself.

Our results show that the tendency to avoid information varies substantially across individuals, but not along any of the standard demographics one might have assumed (e.g. education). This may have especially important implications for disseminating information and raising awareness. Governments (and private actors) currently apply such campaigns broadly based on the assumption that individuals are receptive to the information. However, the impact of information campaigns may (predictably) differ based on people's tendency to avoid information. The expectation of a hedonic cost might motivate some to sidestep such policy efforts. For example, financial literacy interventions, have been found to have small impacts on behavior overall (e.g., Fernandes, Lynch, and Netemeyer 2014). However, effects may be greater for those who are receptive to information and willing to learn even when doing so may be painful. Similarly, the presence of calorie labels does not always help consumers make healthier choices (Elbel et al. 2009). Information disclosure may nonetheless be effective for the subset of consumers who are not predisposed to avoid potentially unpleasant health-related information.

Given the welfare implications of avoidance, people's preferences for information ought to be accounted for when designing interventions to help reduce an unwanted behavior (e.g., smoking cessation) or increase uptake of actions with positive outcome (e.g., more annual physicals). Studies in health behavioral phenotyping have begun to personalize care based on behavioral trends and prior responses to health interventions (Jethwani, Kvedar, and Kvedar 2010). Automated algorithms in the form of robo-advisors now guide the information that is delivered to consumers based on balance and prior investing experience (e.g., Betterment and Wealthfront). Personalized interventions are considered promising in drug development (Ginsburg and McCarthy 2001; Schork 2015; Swan 2009); similarly, personalized messaging campaigns may make informational campaigns more effective. Knowing who is likely to engage with certain kinds of information could improve the effectiveness of informational campaigns and avoid exposing people to information they would be better off not obtaining (in terms of their belief utility) and are unlikely to act on. Information seekers and avoiders may benefit from different messaging, much like extremely risk-averse investors may desire different products than do those who are more tolerant of volatility.

With the recent emergence of information avoidance as a central topic in economics and other disciplines, measuring information preferences in laboratory and field experiments may become as important as measuring risk and time preferences. We hope that the availability of a valid and reliable scale to measure individual differences will prove useful to a diverse set of researchers.

2 The Hidden Cost of Soft Paternalism

2.1 Introduction

In an effort to reduce carbon emissions, governments have traditionally relied on mandates and financial incentives. Existing policies include fuel efficiency standards for cars, taxes on fuels, and subsidies for adopting renewable energy sources. More recently, drawing on insights from psychology, individuals in the United States have begun receiving power bills that compare their personal energy use to that of their neighbors, which has been shown to decrease energy consumption (Allcott 2011). Such psychological interventions, or "nudges," (Thaler and Sunstein 2008) are often costless to implement, but have a measurable impact on behavior. For example, when a green energy plan is the default, more people choose to stay enrolled it in than when they have to explicitly opt-in (Hedlin and Sunstein 2016). Because they do not impose a direct cost or rely on mandates, nudges often find broader support among the public and among policymakers (Hall et al. 2018), especially among those who are inherently skeptical of regulation and mandates (Thaler and Sunstein 2003; Camerer et al. 2003b). Nudges are especially appealing because they are inexpensive for governments to implement (Benartzi et al. 2017), do not limit the autonomy of those who choose differently for good reason (Johnson and Goldstein 2003; Madrian and Shea 2001) and in some domains have had greater impact than standard economic interventions (Bergman and Rogers 2017; Patel et al. 2016). Ultimately, nudges promise to address major policy problems in a variety of domains without the costs associated with standard economic policies (Allcott and Rogers 2014; Yoeli et al. 2017; Loewenstein and Chater 2017a; Marteau et al. 2011; Levitt et al. 2016; Madrian et al. 2017; Thaler and Benartzi 2004; Rogers and Feller, n.d.).

However, nudges may not always be as effective or costless as believed. For example, firms could in turn nudge consumers toward less environmental (and more profitable) products (Sunstein 2017), Nudges also fail to overcome economic incentives that primarily drive carbon emissions. To the extent that nudges do have an effect in the environmental context, influencing one aspect of behavior may give people moral license to offset their behavior elsewhere (Meijers et al. 2015). For example, residents who were successfully nudged to decrease their water consumption increased their use of electricity compared to control households (Tiefenbeck et al. 2013). Compounding this effect, a sense of making (a little) progress on reducing emissions can undermine motivation to do more (Fishbach, Dhar, and Zhang 2006). Indeed, recent findings suggest that merely remembering having taken more actions to reduce one's own energy consumption leads to less support for government action on climate change (Werfel 2017). People may further feel satisfied that a single action to tackle a problem is sufficient and are hesitant to implement multiple interventions (Weber 1997; Hansen, Marx, and Weber 2004). Encouraging a particular action can therefore have substantial negative behavioral spillover (Nash et al. 2017; Thøgersen 1999; Truelove et al. 2014). To evaluate the efficacy of a policy intervention, such behavioral spillovers – both positive and negative – must be accounted for (Dolan and Galizzi 2015).

We propose another, more pernicious cost: support for an environmental nudge may crowd-out support for another (more effective, but costlier) environmental policy. Policymakers who want to be seen as taking action on climate change, and voters who are concerned about the costs of heavy-handed emission reduction policies, such as carbon taxes, may rely on the nudge as an alternative rather than the complementary role that nudges ideally serve. The actual or potential implementation of a nudge may thereby reduce the likelihood that an alternative policy with a bigger effect will be implemented. This subtle downside of nudges has received attention in commentaries (Loewenstein and Ubel 2010; Loewenstein and Chater 2017b), and we provide the first experimental evidence that it indeed occurs with policies aimed at reducing carbon emissions.

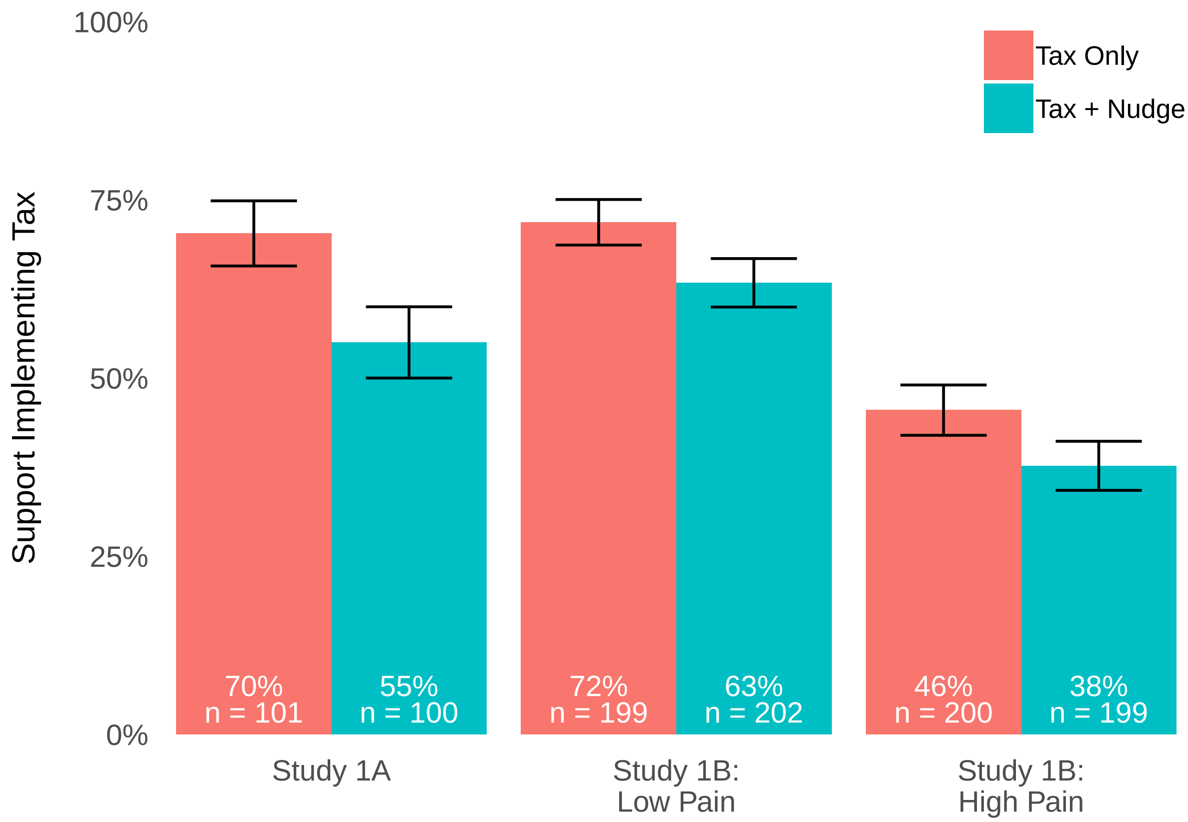
The experiments we report had four main objectives. First, we wanted to see whether allowing participants to implement a green energy nudge in addition to a carbon tax would diminish support for the tax. While passing a nudge to reduce carbon emissions may be preferable to taking no action, doing so may be costly if it undermines, rather than complements, more effective interventions. Second, we tested whether the crowding-out effect applies only to nudges that are also aimed at emission reduction or whether nudges undermine support for economic incentives generally. If our explanation of the nudge as a perceived substitute holds, we should only observe the effect when the nudge and the tax target the same issue. Third, we investigated whether merely manipulating the order in which the two policies are presented would reduce support for the tax. That lets us rule out that our effect is driven by a change in the number of policies available for implementation. This also allowed us to test whether crowding-out occurs in the opposite direction: that is, whether tax policies undermine support for nudges, in a test of the single-action bias applied to environmental policy implementation (Weber 1997; Hansen, Marx, and Weber 2004). Such a bias would predict that the tax also crowds-out support for the nudge, whereas our explanation predicts no crowding-out in that direction. To address concerns of generalizability, particularly to groups who may have different incentives (e.g., because they have some control over which policies get implemented) we replicate the third study with graduates of a policy school, about half of whom had experience influencing public policy. Finally, in a fourth study, we explore means of eliminating crowding-out that also shed light on the mechanism underlying the effect. In all our studies, we further examined whether crowding-out is more pronounced for some demographic groups, for example those less supportive of government mandates, those who believe the carbon tax to be less effective, or those who are more skeptical of the existence of climate change. We find no systematic differences and report these additional results in the Supplementary Information.

2.2 Experimental Results

2.2.1 Study 1

In Studies 1A and 1B, we introduced participants (1A: N = 201, 49.75% female, mean age 34.7; 1B: N = 800, 56.12% female, mean age 35.69) to a nudge defaulting residential consumers into a renewable energy plan (Sunstein 2016) and a $40 per ton carbon tax (Interagency Working Group and others 2013). Participants were randomly assigned to a decision to implement the tax vs. doing nothing (single implementation), or a decision to implement the tax, the nudge, both, or neither

Figure 4: Introducing a green energy nudge into the choice set crowds-out support for the carbon tax (Study 1A). We replicate the findings in Study 1B and show it holds even among those who would support a more painfully framed carbon tax. Error bars show ± one standard error.



(joint implementation). The difference in support for the carbon tax will serve as our measure of crowding-out. In Study 1B, we added a second dimension in which we framed the tax as more painful, making salient higher costs for individuals. This additional manipulation allows us to see whether crowding-out of support occurs also among those who would be willing to accept a more painful policy otherwise (i.e., those who are more committed to taking action).

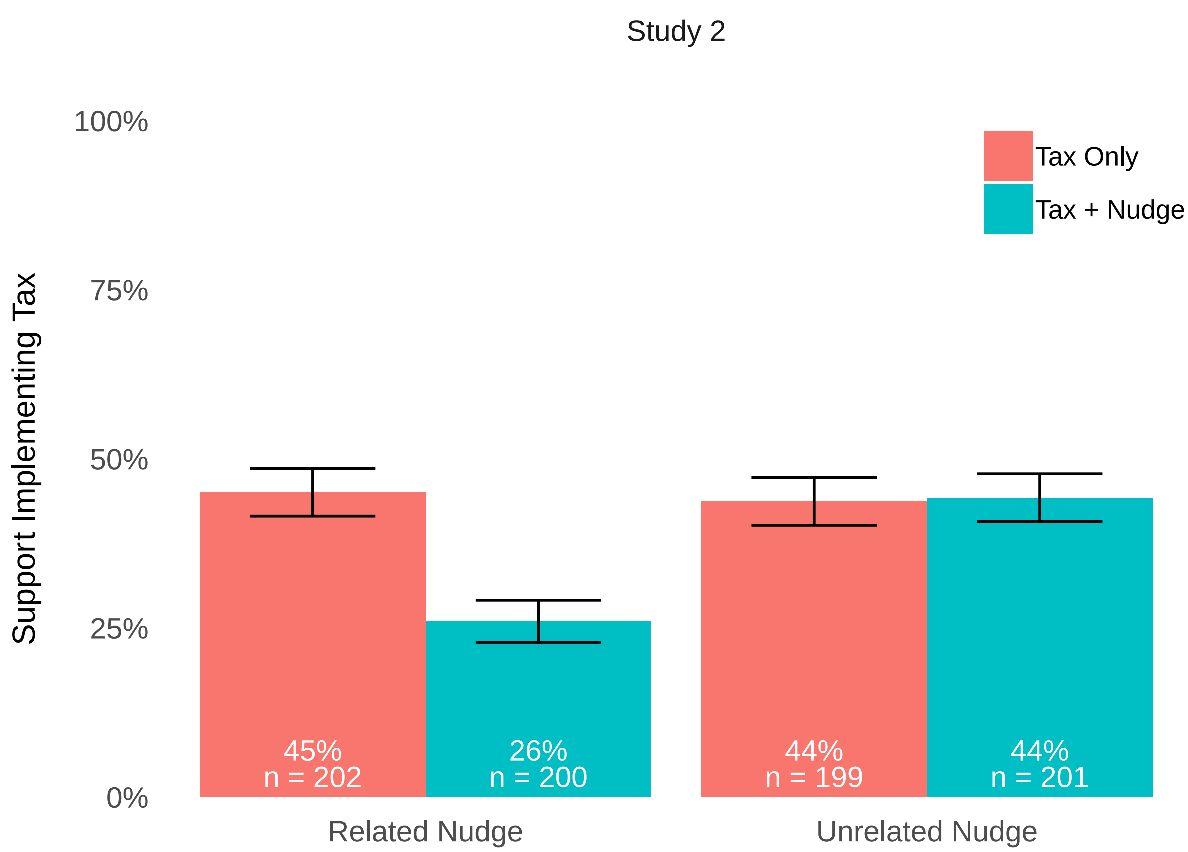
In Study 1A, we find that the carbon tax was perceived as more painful than the green energy nudge (M_{tax} = 2.61, SD_{tax} = 1.11 vs. M_{nudge} = 1.62, SD_{nudge} = 0.85, $t(200) = -11.78$, $p < .001$, using a paired two-sample t-test). For this and all other statistical analyses, we report two-sided test statistics. Unexpectedly, however, we find no difference in the expected efficacy of the two policies (M_{tax} = 3.16, SD_{tax} = 1.19 and M_{nudge} = 3.23, SD_{nudge} = 1.01, $t(200) = 0.86$, $p = .390$), even though the tax is, in fact, dramatically more effective at reducing carbon emissions. Figure 4 shows the level of support for the carbon tax (or both the tax and the nudge) for both studies. When participants can implement only the carbon tax, but not the nudge, we find fairly high support for the tax (70.3%). However, when a green energy nudge becomes available to implement, many fewer respondents favor implementing either the tax only or both (55%, $t(197.23) = 2.26$, $p = .025$). The “low-pain” framing of the tax in Study 1B replicates these results, with 71.86% and 63.37% supporting the tax ($t(397.85) = 1.82$, $p = .069$). When we frame the tax as more painful, we observe a decrease in support for the tax compared to the low-pain framing, both without the nudge (71.86% vs. 45.5%, $t(393.33) = 5.54$, $p < .001$) and with the nudge (63.37% vs. 37.69%, $t(398.83) = 5.31$, $p < .001$). However, there is no significant interaction between the framing and the effect of introducing a nudge (regression analyses for this and subsequent studies shown in the Supplementary Information).

We find that introducing a green energy nudge crowds-out support for a more effective carbon tax. Moreover, this effect does not merely displace support among those marginally willing to implement a policy: even those who would otherwise be open to accepting a painful-sounding policy are enticed by the less painful nudge.

2.2.2 Study 2

To test whether nudges undermine support for standard policies in general, Study 2 presents participants in one condition with a nudge outside the climate change domain. We predicted that only the green energy nudge would crowd-out support for the carbon tax, as the other does not serve as a potential substitute. This design also addresses the concern that participants may merely be reluctant to implement two policies (Weber 1997; Hansen, Marx, and Weber 2004) or that the results are merely an artifact of our particular experimental design. Participants (N = 802, 53.74% female, mean age 35.54) read about the green energy nudge or instead learned about a nudge defaulting employees into an employer-sponsored pension plan. They then all read about the painfully framed

Figure 5: When the nudge tackles a different policy problem than the tax, we observe no crowding-out. When the policies are related, we replicate the findings from Studies 1A and 1B. Error bars show ± one standard error.



carbon tax policy from Study 1B. In both conditions, half of all participants again could only implement the carbon tax (vs. nothing) and the other half could implement the tax, the nudge, both, or nothing.

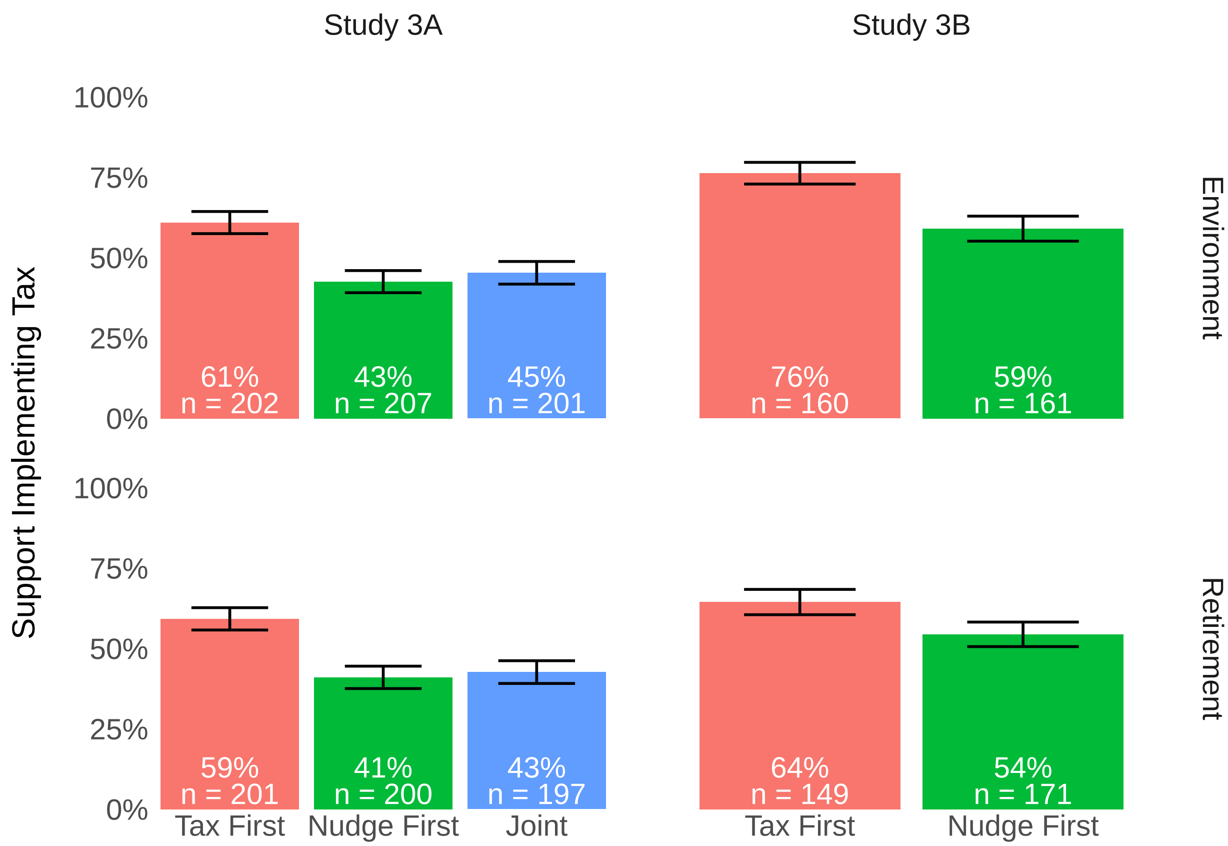
As before, we find high support for the carbon tax when participants could implement the carbon tax but not the green energy nudge (45.05% of participants supported the tax when they read about the green energy nudge and 43.72% when they read about the retirement savings nudge, $t(398.94) = -0.27, p = .789$). When we introduced the relevant green energy nudge into the choice set, we replicate our prior result and find a reduction in support for the tax (26%, $t(394.74) = 4.06$, $p < .001$). When the nudge shown is in the retirement savings domain, however, we find no similar displacement (44.28%, $t(397.97) = -0.11$, $p = .910$). Regressions (see Supplementary Information) show that there is indeed a significant interaction: crowding-out occurs only when the two policies are both in the environmental domain ($p < 0.001$).

2.2.3 Study 3

Our next pair of studies looks at whether merely presenting a nudge prior to a tax can lead to the same crowding-out. In addition to an environmental policy pair (a green energy nudge and a carbon tax), we also introduce a pair of policies aimed at promoting retirement savings (a 401(k) default nudge and an expansion of the social security tax). This allows us to test the robustness of the effect to a variety of policy pairs. If crowding-out is indeed more general, as we hypothesize, then it is likely to affect other interventions aimed at reducing carbon emissions and promoting pro-environmental behavior. Moreover, we extend our results to a novel participant pool: alumni of a public policy school, many of whom report having experience in policymaking. This allows us to generalize our finding to a sample of experts who have more informed policy views and consequently might be less susceptible to being influenced by the availability of a nudge.

In Study 3A, we recruited participants (N = 1208, 55.96% female, mean age 36.19) and randomly assigned them to conditions in a between-subjects $2x3$ design. Participants were presented either

Figure 6: When participants make sequential implementation decisions, choosing whether to implement the nudge first reduces support for the tax similarly to the joint implementation decision previously (Study 3A). Our findings replicate with a sample trained in public policy (Study 3B) and occur in both the environment and retirement domains (3A and 3B). Error bars show ± one standard error.



with the green energy nudge and the low-pain carbon tax from Study 1 ("Environment") or a 401(k) savings default nudge and an expansion of social security ("Retirement"). On the second dimension, we varied whether participants first learned about the nudge and got the choice to implement it before doing the same with the tax ("Nudge First"); vice versa ("Tax First"); or learned about both policies before having the opportunity to implement either one, both, or neither as in our previous studies ("Joint Implementation"). We preregistered the study design, hypotheses, and analyses on AsPredicted.org.

The left panel of Figure 6 shows the fraction of participants who supported implementing the tax in the two domains. Beginning with the environmental domain, we observe that support is greatest when the decision to implement the tax was offered first (60.89%) and drops when participants could first implement the nudge (42.51%, $t(406.95) = 3.77$, $p < .001$) and when they made both decisions simultaneously (45.27%, $t(400.75) = 3.17$, $p = .002$). We observe no difference between the "Nudge First" and the "Joint Implementation" decision ($t(405.46) = 0.56$, $p = .575$), suggesting that our findings are not merely a result of asking about two policies at once. The results are similar in the retirement domain, suggesting this is not specific to the selected pair of policies either. Again, we see the most support for expanding social security when the choice was offered first (59.2%), compared to when the nudge could be implemented first (41%, $t(398.99) = 3.70$, $p < .001$) or when the decision for the two policies was made jointly (42.64%, $t(395.72) = 3.34$, $p = .001$). Here, too, we observe no difference between the last two conditions ($t(394.83) = 0.33$, $p = .741$).

To rule out that respondents were merely averse to implementing two policies on one issue (single-action bias), we look at support for the nudge and find no change across experimental conditions (Environment: 75% support, $F(2, 607) = 0.45$, $MSE = 0.12$, $p = .640$; Retirement: 75% support, $F(2, 595) = 0.59$, $MSE = 0.19$, $p = .553$).

For Study 3B, we recruited alumni of a public policy school (N = 641, 46.8% female, mean age 46.9). The sample is highly educated (87.99% have a Master's Degree and 11.86% have a PhD) and 54.13% report holding or having held a position in which they influenced public policy. We preregistered the study design, hypotheses, and analyses on AsPredicted.org.

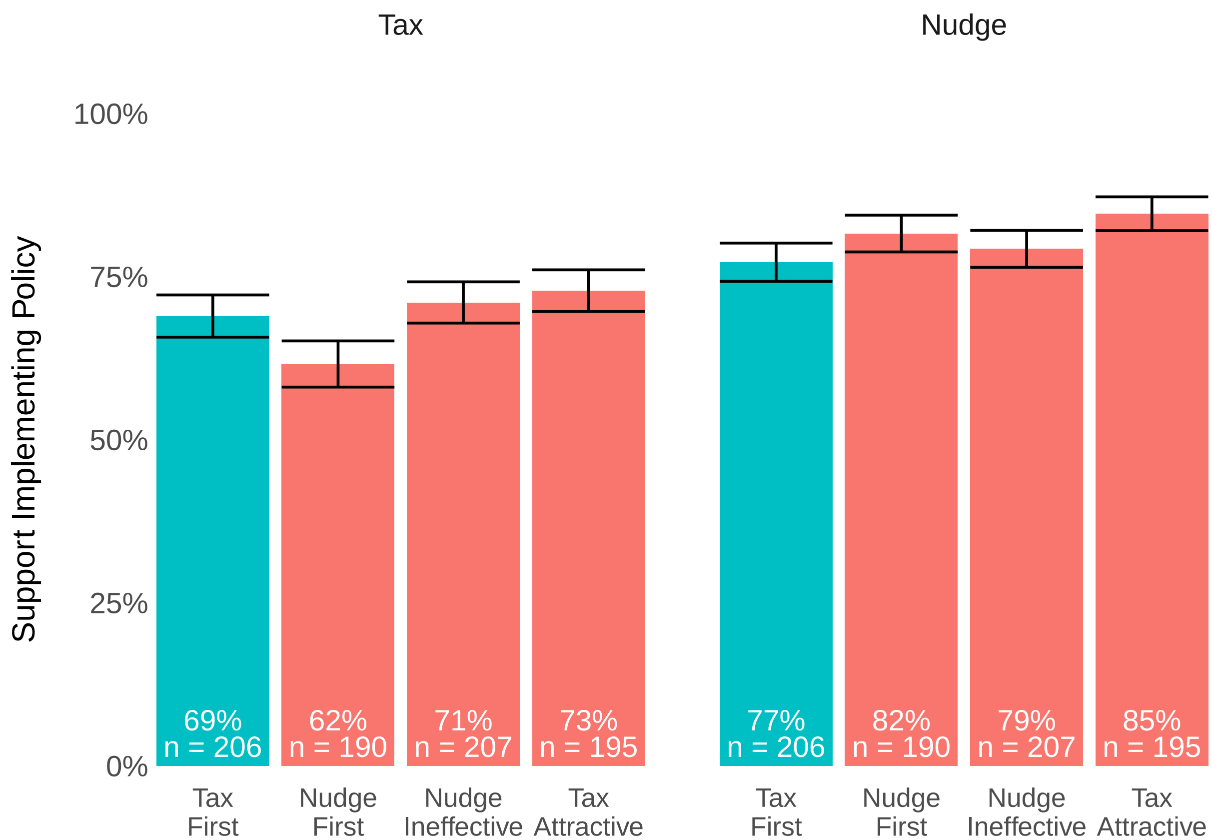
As with our Mechanical Turk sample, we find that respondents in our study thought the carbon tax was more painful than the green energy nudge (M_{tax} = 2.69, SD_{tax} = 1 vs. M_{nudge} = 1.59, SD_{nudge} = 0.83, $t(620.43) = -15.10$, $p < .001$) and the social security expansion more painful than a 401(k) nudge (M_{tax} = 2.44, SD_{tax} = 1.02, vs. M_{nudge} = 1.53, SD_{nudge} = 0.83, $t(610.57) = -12.40$, $p < .001$). Curiously, we find that this sample believed the carbon tax to be even less effective than the green energy nudge (M_{tax} = 2.96, SD_{tax} = 1.05, vs. M_{nudge} = 3.12, SD_{nudge} = 0.94, $t(631.90) = 2.06$, $p = .040$) and that they believed an expansion of social security to be less effective than the 401(k) savings nudge in promoting retirement savings (M_{tax} = 2.94, SD_{tax} = 1.28, vs. M_{nudge} = 3.27, SD_{nudge} = 0.98, $t(596.54) = 3.74$, $p < .001$).

The right panel of Figure 6 shows the fraction of participants who supported implementing either the carbon tax ("Environment") or expanding the social security tax ("Retirement"). When the tax was presented first, we see greater support for its implementation in both the environmental domain (76.25% vs 59.01%, $t(313.08) = -3.35$, $p = .001$) and in the retirement domain (64.43% vs 54.39%, $t(314.91) = -1.83$, $p = .068$), although it only reaches the usual level of significance in the former. If we look separately at the subset of participants who report applied policymaking experience, we observe the same findings as in the full sample of participants reported here (see Supplementary Information).

2.2.4 Study 4

Finally, in Study 4 we look at two potential mechanisms that could lead to crowding-out. People might believe the nudge to be more effective than it is, or they might be motivated to avoid a painful tax. In the "Tax First" condition, we ask participants about their support for the carbon tax, followed by their support for the green energy nudge. The three remaining conditions ask about the nudge first, either providing no additional information ("Nudge First"), highlighting that only a small proportion of people are nudged by a green energy default and most carbon emissions are due to other sources ("Nudge Ineffective"), or making the tax less costly by highlighting that revenue can be used to offset other taxes ("Tax Attractive"). Figure 7 shows, in the left panel, support for the carbon tax, which declines from 69% to 62% when we change the ordering of the policies from tax first to nudge first, although this difference is not statistically significant ($t(387.36) = 1.53$, $p = .126$). Asking about the nudge first, but highlighting its effect size, increases support compared to omitting that information (71%, $t(385.69) = -1.99$, $p = .047$); as does highlighting that other taxes could be lowered to offset the additional costs (73%, $t(378.00) = -2.36$, $p = .019$). As before, we observe no difference in support for the nudge across conditions ($F(3, 794) = 1.30$, $MSE = 0.16$,

Figure 7: Support for implementing the carbon tax and the green energy default nudge by condition. Being explicit about the nudge's effect size eliminates crowding-out on the tax, but does not diminish support for the nudge. The same holds when increasing the attractiveness of the nudge. Error bars show $\pm$ one standard error.



$p = .272$). Notably, adding information about the effect size to the nudge does not diminish support compared to omitting that information (79% and 79%, respectively, $t(394.35) = 0.59$, $p = .556$)

2.3 Discussion

Across our six studies, we found that support for a carbon tax can decline when a green energy nudge, a much less effective but also less painful policy, is introduced. We find no consistent heterogeneous treatment effects that would suggest such crowding-out is more pronounced for those more opposed to government intervention, less certain that climate change is occurring, or who believe the nudge to be more effective than the other policy (see Supplementary Information).

What accounts for the crowding-out effect of nudges on more substantive policies? We believe that the account which best fits the data just presented is that while people are, generally, concerned about societal problems such as climate change, they may not be willing to incur large costs to achieve a solution. Introducing a nudge raises the possibility of a low-cost solution, and people may then engage in motivated reasoning to exaggerate the effectiveness of this tempting alternative and focus solely on the (economic) cost-effectiveness of the intervention, disregarding the magnitude of the policy's true environmental impact. This may explain why, on average, participants thought the nudge was as or more effective at reducing pollution than the carbon tax. Moreover, even those who (rightfully) believed the carbon tax to be more effective than the green energy nudge were equally discouraged from implementing the tax when a nudge became available, suggesting that crowding-out is not merely the result of incorrect perception of relative effectiveness. When these perceptions are corrected at the time of decision, however, motivated reasoning becomes more difficult and indeed crowding-out disappears.

The effects documented in these studies are, very likely, more general and not restricted to the domains of climate change and retirement savings, or to nudges and more substantive policies. The possibility of introducing an information campaign to educate people about how they can personally cut their emissions, for example, might be perceived as even less invasive than a green energy nudge and crowd-out support for the latter (Davidai and Shafir 2018).

Our studies presented respondents with hypothetical decisions, and elicited their support for policies using subjective measures. Although we find our effect even among those with training in public policy and among a subset of responders with experience in policymaking, we cannot know whether the effects documented in our research will ultimately have an impact in the legislative or regulatory processes involved in crafting and implementing environmental regulation. However, as Reisch and Sunstein write, "public officials are inevitably responsive to what people think,"(p. 311 Reisch and Sunstein 2016) and to that extent, reduced support for a carbon tax among voters may indeed make it less likely to be implemented. We further show in Study S1 that crowding-out occurs not only in an implementation decision, but also in reported favorability of a policy: When respondents were first asked how favorably they viewed a green energy nudge, they subsequently viewed the carbon tax less favorably. Such a question may naturally be asked in opinion polls and shape the agenda of public officials.

In an ideal world, we could tackle climate change with both nudges and more heavy-handed interventions. However, as our results suggest, an effort to actively promote all available tools may have the unintended consequence of reducing the likelihood that the most effective policies will be implemented. Downplaying the impact of nudges appears to resolve this issue without undermining support for nudges themselves.

2.4 Methods

2.4.1 Studies 1A and 1B

Experimental materials for all studies are presented in the online appendix.

In Study 1A, we recruited 201 participants via Amazon Mechanical Turk for a "Study on Decision Making" that was expected to take 5 minutes to complete and offered a fixed payment of 50 cents. We began by asking participants whether they agreed that global average temperatures had been increasing over the past 50 years (independent of the cause) and, for all but those who strongly disagreed, we asked whether they believed human activity to be the primary cause.

We next presented them with a brief description explaining default nudges. Participants were informed that mandating a default can lead to an option being chosen more frequently, without prohibiting people from choosing differently. We then introduced them to the green energy nudge, using the description from previous work examining attitudes toward this nudge.(Sunstein 2016) Participants then rated how effective they believed the policy to be at reducing pollution and carbon emission on a 5-point Likert scale from "Not effective at all" to "Extremely effective." On the same screen, they also rated how painful they thought the policy would be for someone like them, on a

5-point Likert scale from “Not painful at all” to “Extremely painful”. Next, we introduced them to a \$40/ton carbon tax that would be levied on companies and individuals. This tax was described as capturing the economic costs of carbon emissions.(Interagency Working Group and others 2013) Participants then evaluated this policy on the same two dimensions as the nudge.

We then randomly assigned participants to one of two conditions. In the “Tax Only” condition, participants imagining themselves in the role of a policymaker were asked to choose whether to implement the carbon tax policy. They were told that the alternative, if the policy is not implemented, is that no other policy would be implemented. They could then choose to implement the tax or not implement the tax. In the “Tax and Nudge” condition, participants were also asked to imagine themselves as policymakers. However, they had more choices: they could implement the tax, the nudge, both, or neither. They were also told that if neither of the policies were implemented, no other policy would be passed in their place. The survey concluded with basic demographic questions: gender, age, ethnicity, education, and political orientation.

In Study 1B, we recruited 800 new participants via Amazon Mechanical Turk for a five minute study on decision making that paid 50 cents. Participants were randomly assigned to one of four conditions: “Low Pain, Tax Only,” “Low Pain, Tax and Nudge,” “High Pain, Tax Only,” and “High Pain, Tax and Nudge.” The two “Low Pain” conditions were identical to the two conditions reported in Study 1A and are direct replications. The two “High Pain” conditions differed in that they included additional descriptive information about the carbon tax conveying the cost to consumers. In particular, for those in the high cost condition, we point out that this policy would *substantially* raise the price not just on transportation, but also on heating and air conditioning, on electricity, and on other goods and activities.

2.4.2 Study 2

We recruited 802 participants from Amazon Mechanical Turk for a five minute study on decision making in exchange for a fixed payment of 50 cents. When participants entered the survey, they were informed that they would be asked to evaluate two policies aimed at addressing longstanding problems. Half the participants were then randomly assigned to the “Related Nudge” condition. They first read about the green energy nudge from Study 1A, in which the government would require large electricity providers to enroll consumers into plans with environmentally friendly energy suppliers; but noting that consumers could opt-out if they wished. We then asked them to evaluate how effective the policy would be if it were implemented and how painful it would be for someone like them. Both responses were reported on a 5-point Likert scale from “Not effective (painful) at all” to “Extremely effective (painful).” The other participants were assigned to the “Unrelated Nudge” condition and instead read about a retirement nudge in which large employers would be required to enroll employees into a pension plan, but would allow them to opt out if they wished.(Sunstein 2016) They, too, evaluated these policies according to their effectiveness and painfulness.

Next, all participants read about the same carbon tax policy: a \$40/ton tax. We used the “High Pain” phrasing from Study 1B, noting that the policy would substantially raise the price on transportation, on heating and air conditioning, on electricity, and on other goods and activities. Participants evaluated this policy, too, on the effectiveness and painfulness dimensions. Finally, participants were asked to imagine themselves as a policymaker and were given a decision to implement a policy. In the “Tax Only” condition, we asked them if they wanted to implement the carbon tax. We noted that no other policy would be passed if they decided not to do so. In the “Tax and Nudge” condition, we offered them four choices: implement the carbon tax only, implement the nudge only, implement both the tax and the nudge, or implement neither the tax nor the nudge. Which nudge participants got to implement depended on which one they had randomly been assigned to read about: either the green energy nudge or the retirement savings nudge.

2.4.3 Study 3A

In Study 3A, we recruited 1208 participants and randomly assigned them to conditions in a $2x3$ between-subjects design. On the first dimension, we vary whether participants face policies in the domain of climate change ("Environment") or retirement savings ("Retirement"). In all conditions, participants were asked to evaluate on 5-point Likert scales how effective and how painful each of two policies would be (from "not at all" to "very").

In the "Environment" domain, we began by introducing the threat of climate change and told participants that they would be asked to evaluate two policies aimed at combating it. The two policies were the green energy nudge (identical to previous studies) and a carbon tax. For the tax, we used the framing from Study 1A and, identically, the "Low Pain" condition of Study 1B. In the "Tax First" condition, participants began by evaluating the carbon tax, then made a decision about whether or not to implement it. They then were presented with the green energy nudge and, after evaluating it, were asked about whether or not they would implement that policy. In the "Nudge First" condition, we reversed the order: participants first decided whether to implement the nudge, then made the decision about the tax. Finally, the "Joint Implementation" condition matches our previous design: participants first read about the nudge, then about the tax, and only at the end got to decide which of the policies, if any, to implement.

In the "Retirement" domain, we introduced the problem of undersaving and told participants that they would be asked to evaluate two policies that may increase the income people have available in retirement. In the "Tax First" condition, participants evaluated an expansion of social security. The program would increase contribution rates for employees and employers, but would also increase benefits and eliminate uncertainties about the availability of future benefits. They then read about a 401(k) savings nudge, in which large employers would be required to enroll workers into a retirement plan and contribute 8% of their income by default. Employees would have the option to change the savings rate or opt-out entirely. In the "Nudge First" condition, we reversed the order in which the two policies were presented: participants first read about (and got to implement) the

401(k) savings nudge, then read about (and got to implement) the expansion of the social security program. Finally, in the "Joint Implementation" decision, participants read about and evaluated first the 401(k) savings nudge, then the expansion of social security, and only at the end had the option to implement one of the policies, both, or neither.

We concluded the survey with basic demographic questions from the previous studies (gender, age, ethnicity, education, and political affiliation). The experimental design, sample size, hypotheses, and planned analyses were preregistered on AsPredicted.org #5424. A link to this preregistration report will be available in the publication version of this paper.

2.4.4 Study 3B

For Study 3B, we recruit a sample of participants with training in public policy. We contacted all 4,455 alumni of the Heinz College of Public Policy at Carnegie Mellon University whose email addresses were on file with the alumni office. Of those contacted, 835 clicked on the link in the email and 641 completed the survey. The average age was 46.9 and 53.2% were male. The sample was highly educated: 87.99% had a Master's Degree, 11.86% a doctorate, and 41.34% had taken a graduate-level behavioral economics class.

Our respondents are also actively involved in the shaping of public policy: 54.13% stated that their current or past roles involved public policy either directly or indirectly. We refer to this subgroup as "policymakers" and perform all our analyses separately on them as an additional robustness check. Within this sample of policymakers, the average age was 45.75 and 58.79% were male. Among policymakers, 83.57% obtained Master's Degrees, 16.14% obtained doctorates, and 42.65% had taken a graduate-level behavioral economics class.

The design follows closely that of Study 3A, and aimed to replicate its findings with a more informed sample. However, as we did not want to risk being underpowered on our main comparison of interest, we dropped the "Joint Implementation" condition, leaving the "Tax First" and "Nudge First" conditions. Moreover, participants saw a tax and a nudge in each of the two policy domains

(Environment and Retirement). We randomized between-subjects the order in which the policies and the domains were presented, using a Latin Square design. Those who first saw the tax and then the nudge in the first domain subsequently first saw the nudge, followed by the tax, in the second domain. This allows us to test for spillover across domains, where the decision to implement policies in the retirement domain might allow them to anticipate the policies they will face in the environmental domain, and vice versa. Absent spillover, we could collapse across the domain ordering and increase our sample size. In addition to analyzing the full sample of respondents, we perform separate analyses on the subset of respondents who reported being directly involved in shaping public policy.

In both the domains of climate change and retirement savings, participants read about two kinds of policies: taxes and default nudges. In the environmental domain, the standard policy imposes a carbon tax on companies based on how much emissions they create, which in turn raises the price of goods. The nudge consists of a mandate on large energy providers to automatically enroll consumers into a green power plan, though consumers can elect to opt-out if they wish. The standard economic policy for retirement savings consisted of an increase in the social security tax for both employees and employers, along with a commensurate increase in social security benefits. The corresponding nudge was a mandate for employers to enroll workers into contributing 8% of their salary into a 401(k) plan, but allowing workers to opt-out or change their allocation.

After reading about each of the four policies, participants were asked how effective the policy would be at increasing retirement savings or reducing pollution and mitigating CO_2 emissions, and whether they wished to implement the policy. They then proceeded to the next policy. The survey concluded with demographic questions. The study and analyses, including the test for spillover across the two domains and the subgroup analysis for policymakers, were preregistered on AsPredicted.org #5624. The preregistration report mistakenly notes that we had already collected some data, which is inaccurate; the timestamp of the preregistration report precedes the distribution of emails by two days.

2.4.5 Study 4

We recruit participants from Amazon Mechanical Turk for a five minute study in exchange for a 50 cent fixed payment. We targeted a sample of 800 people who passed an attention check included at the end of the experiment. After recruiting 954 participants, we ended up with 798 participants who passed. The sample size, exclusion criteria, along with the following experimental design, hypotheses, and analyses were pre-registered on AsPredicted #15694.

We began by introducing all participants to the threat of climate change and the effect of pollution on premature deaths. Next, participants read that they would be asked to evaluate two policies that governments might considered to combat pollution and global climate change. We then randomly assigned them to one of four between-subjects conditions. In the "Tax First" condition, participants read about the $40/ton carbon tax from the previous study and rated it on effectiveness and painfulness (again on a 5-point Likert scale from "Not at all" to "Very"). On the next screen, they were then asked if they would vote to implement the carbon tax (using the low-pain framing from Study 1A and Studies 3A and 3B). Independent of their answer, they were then presented with the green energy nudge, also identical to the previous experiments. Similarly, they evaluated it on effectiveness and painfulness and stated whether they would vote to implement it.

The experimental survey then asked a series of demographic questions: gender, age, ethnicity, education, political affiliation and political orientation. We also asked participants how they believe their carbon emissions compared to the average household (more, less, or the same) and which one of four statements most closely reflected their views on climate change. They could express that climate change was primarily caused by human activity and governments should take measures to reduce emissions; primarily caused by human activity, but actions to reduce should be left to individuals; primarily caused by natural factors, but governments should take measures to reduce emissions; and primarily caused by natural factors and governments should not take measures to reduce emissions. We did not preregister any hypotheses related to these questions, but collected the responses for descriptive purposes.

Participants in the “Nudge First” condition faced a survey that was identical except in that they first evaluated the green energy nudge, then the carbon tax. The remaining two conditions, “Nudge Ineffective” and “Tax Attractive,” followed the same order as the “Nudge First” condition. However, those two conditions received additional information. In the “Nudge Ineffective” condition, we told participants that green energy nudges have been found to only shift a fraction of the population toward green energy, that residential electricity use is responsible for only a small part of carbon emissions, and that the policy would hence have very little impact on emissions. This information was truthful.(Hedlin and Sunstein 2016)

In the “Tax Attractive” condition, we instead provided more information about the carbon tax. We told participants that British Columbia had implemented a similar tax and uses part of the revenue to lower income taxes. We further highlighted that revenue could be used to lower other taxes and fund projects we thought would be appealing to participants. Moreover, we noted that a previous ballot initiative in Washington would have returned the revenue to residents, which would lead households that emit less carbon than average to receive a greater rebate than what they paid in taxes.

The survey concluded with an attention check. We first showed participants an image of a bell pepper and asked them what they saw in the image. Anyone whose response included the word “pepper” was marked as having passed the attention check. We then asked participants to write the date “08/06/2018” in words. We treat anyone whose response included the word “August” as having passed the attention check.

2.5 Ethical Approval

For all studies, we obtained ethical approval from the internal review board at Carnegie Mellon University and complied with all relevant ethical regulations for research with human participants. None of our studies involved deception.

2.6 Data Availability

The raw data from all our experiments and statistical code for all analyses and figures reported in the paper and the supplementary analyses will be available via Github and as an R library on CRAN following publication of the paper.

3 Warning: You Are About to be Nudged

3.1 Introduction

Nudging people toward particular decisions by presenting one option as the default can influence important life choices. If a form enrolls employees in retirement savings plans by default unless they opt out, people are much more likely to contribute to the plan (Madrian and Shea 2001). Likewise, making organ donation the default option rather than just an opt-in choice dramatically increases rates of donation (Johnson and Goldstein 2003). The same principle holds for other major decisions, including choices about purchasing insurance and taking steps to protect personal data (Johnson et al. 1993; Acquisti, John, and Loewenstein 2013). Decisions about end-of-life medical care are similarly susceptible to the effects of defaults. Two studies found that default options had powerful effects on the end-of life choices of participants preparing hypothetical advance directives. One involved student respondents, and the other involved elderly outpatients (Kressel and Chapman 2007; Kressel, Chapman, and Leventhal 2007). In a more recent study, defaults also proved robust when seriously ill patients completed real advance directives (Halpern et al. 2013).

The use of such defaults or other behavioral nudges (Thaler and Sunstein 2008) has raised serious ethical concerns, however. The House of Lords Behaviour Change report produced in the United Kingdom in 2011 contains one of the most significant critiques (House of Lords 2011). It argued that the "extent to which an intervention is covert" should be one of the main criteria for judging if a nudge is defensible. The report considered two ways to disclose default interventions: directly or by ensuring that a perceptive person could discern a nudge is in play. While acknowledging that the former would be preferable from a purely ethical perspective, the report concluded that the latter should be adequate, "especially as this fuller sort of transparency might limit the effectiveness of the intervention." Philosopher Luc Bovens in "The Ethics of Nudge" noted that default options "typically work best in the dark (Bovens 2009)." Bovens observed the lack of disclosure in a study in which healthy foods were introduced at a school cafeteria with no explanation, prompting students to eat fewer unhealthy foods. The same lack of transparency existed during the rollout of the Save More Tomorrow program, which gave workers the option of precommitting themselves to increase their savings rate as their income rose in the future. Bovens noted,

> If we tell students that the order of the food in the Cafeteria is rearranged for dietary purposes, then the intervention may be less successful. If we explain the endowment effect [the tendency for people to value amenities more when giving them up than when acquiring them] to employees, they may be less inclined to Save More Tomorrow.

When we embarked on our research into the impact of disclosing nudges, we understood that alerting people about defaults could make them feel that they were being manipulated. Social psychology research has found that people tend to resist threats to their freedom to choose, a phenomenon known as psychological reactance (Wortman and Brehm 1975). Thus, it is reasonable to think, as both the House of Lords report and Bovens asserted, that people would deliberately resist the influence of defaults (if informed ahead of time, or preinformed) or try to undo their influence (if told after the fact, or postinformed). Such a reaction to disclosure might well reduce or even eliminate the influence of nudges. But our findings challenge the idea that fuller transparency substantially harms the effectiveness of defaults. If what we found is confirmed in broader contexts, fuller disclosure of a nudge could potentially be achieved with little or no negative impact on the effectiveness of the intervention. That could have significant practical applications for policymakers trying to help people make choices that are in their and society's long-term interests while disclosing the presence of nudges.

3.2 Testing Effects from Disclosing Defaults

We explored the impact of disclosing nudges in a study of individual choices on hypothetical advance directives, documents that enable people to express their preferences for medical treatment for times when they are near death and too ill to express their wishes. Participants completed hypothetical advance directives by stating their overall goals for end-of-life care and their preferences for specific life-prolonging measures such as cardiopulmonary resuscitation and feeding tube insertion. Participants were randomly assigned to receive a version of an advance directive form on which the default options favored either prolonging life or minimizing discomfort. For both defaults, participants were further randomly assigned to be informed about the defaults either before or after completing the form. Next, they were allowed to change their decisions using forms with no defaults included. The design of the study enabled us to assess the effects of participants' awareness of defaults on end-of-life decisionmaking.

We recognize that the hypothetical nature of the advance directive in our study may raise questions about how a similar process would play out in the real world. However, recent research by two of the current authors and their colleagues examined the impact of defaults on real advance directives (Halpern et al. 2013) and obtained results similar to prior work on the topic examining hypothetical choices (Kressel and Chapman 2007; Kressel, Chapman, and Leventhal 2007). All of these studies found that the defaults provided on advance directive forms had a major impact on the final choices reached by respondents. Just as the question of whether defaults could influence the choices made in advance directives was initially tested in hypothetical tasks, we test first in a hypothetical setting whether alerting participants to the default diminishes its impact. To examine the effects of disclosing the presence of defaults, we recruited via e-mail 758 participants (out of 4,872 people contacted) who were either alumni of Carnegie Mellon University or New York Times readers who had consented to be contacted for research. Respondents were not paid for participating. Although not a representative sample of the general population, the 1,027 people who participated included a large proportion of older individuals for whom the issues posed by the study are salient. The mean age for both samples was about 50 years, an age when end-of-life care tends to become more relevant. (Detailed descriptions of the methods and analysis used in this research are published in the Appendix.)

Our sample populations are more educated than the U.S. population as a whole, which reduces the extent to which we can generalize the results to the wider population. However, the study provides information about whether the decisions of a highly educated and presumably commensurately deliberative group are changed by their awareness of being defaulted, that is, having the default options selected for them should they not take action to change them. Prior research has documented larger default effects for individuals of lower socioeconomic status (Madrian and Shea 2001; Haisley et al., n.d.) which suggests that the default effects we observe would likely be larger in a less educated population.

3.3 Obtaining End-of-Life Preferences

Participants completed an online hypothetical advance directive form. First, they were asked to indicate their broad goals for end-of-life care by selecting one of the following options:

- I want my health care providers and agent to pursue treatments that help me to live as long as possible, even if that means I might have more pain or suffering.
- I want my health care providers and agent to pursue treatments that help relieve my pain and suffering, even if that means I might not live as long.
- I do not want to specify one of the above goals. My health care providers and agent may direct the overall goals of my care.

Next, participants expressed their preferences regarding five specific medical life-prolonging interventions. For each question, participants expressed a preference for pursuing the treatment (the prolong option), declining it (the comfort option), or leaving the decision to a family member or other designated person (the no-choice option). The specific interventions included the following:

- cardiopulmonary resuscitation, described as "manual chest compressions performed to restore blood circulation and breathing";
- dialysis (kidney filtration by machine);
- feeding tube insertion, described as "devices used to provide nutrition to patients who cannot swallow, inserted either through the nose and esophagus into the stomach or directly into the

Table 6: Experimental design.

Group 1: Comfort preinformed	**Group 2: Comfort postinformed**	**Group 3: Prolong preinformed**	**Group 4: Prolong postinformed**
Disclosure		Disclosure	
Choice 1 with Comfort default	Choice 1 with Comfort default	Choice 1 with Prolong default	Choice 1 with Prolong default
	Disclosure		Disclosure
Choice 2 with No default	Choice 2 with No default	Choice 2 with No default	Choice 2 with No default

stomach through the belly";

- intensive care unit admission, described as a "hospital unit that provides specialized equipment, services, and monitoring for critically ill patients, such as higher staffing-to-patient ratios and ventilator support"; and
- mechanical ventilator use, described as "machines that assist spontaneous breathing, often using either a mask or a breathing tube."

The advance directive forms that participants completed randomly defaulted them into either accepting or rejecting each of the life-prolonging treatments. Those preinformed about the use of defaults were told before filling out the form; those postinformed learned after completing the form.

One reason that defaults can have an effect is that they are sometimes interpreted as implicit recommendations (Johnson and Goldstein 2003, 2004; Halpern, Ubel, and Asch 2007; McKenzie, Liersch, and Finkelstein 2006). This is unlikely in our study, because both groups were informed that other study participants had been provided with forms populated with an alternative default. This disclosure also rules out the possibility that respondents attached different meanings to opting into or out of the life-extending measures (for example, donating organs is seen as more altruistic in countries in which citizens must opt in to donate than in countries in which citizens must opt out of donation, Davidai, Gilovich, and Ross 2012) or the possibility that the default would be perceived as a social norm (that is, a standard of desirable or common behavior). After completing the advance directive a first time (either with or without being informed about the default at the outset), both groups were then asked to complete the advance directive again, this time with no defaults. Responses to this second elicitation provide a conservative test of the impact of defaults.

Defaults can influence choices if people do not wish to exert effort or are otherwise unmotivated to change their responses. Requiring people to complete a second advance directive substantially reduces marginal switching costs (that is, the additional effort required to switch) when compared with a traditional default structure in which people only have to respond if they want to reject the default. In our two-stage setup, participants have already engaged in the fixed cost (that is, expended the initial effort) of entering a new response, so the marginal cost of changing their response should be lower. The fact that the second advance directive did not include any defaults means that the only effect we captured is a carryover from the defaults participants were given in the first version they completed. In sum, the experiment required participants to make a first set of advance directive decisions in which a default had been indicated and then a second set of decisions in which no default had been indicated.

Participants were randomly assigned into one of four groups in which they were either preinformed or postinformed that they had been assigned either a prolong default or a comfort default for their first choice, as depicted in Table 6. The disclosure on defaults for the preinformed group read as follows:

> The specific focus of this research is on “defaults”—decisions that go into effect if people don’t take actions to do something different. Participants in this research project have been divided into two experimental groups. If you have been assigned to one group, the Advance Directive you complete will have answers to questions checked that will direct health care providers to help relieve pain and suffering even it means not living as long. If you want to choose different options, you will be asked to check off a different option and place your initials beside the different option you select. If you

> have been assigned to the other group, the Advance Directive you complete will have answers to questions checked that will direct health care providers to prolong your life as much as possible, even if it means you may experience greater pain and suffering.

The disclosure for the postinformed group was the same, except that participants in this group were told that that they had been defaulted rather than would be defaulted.

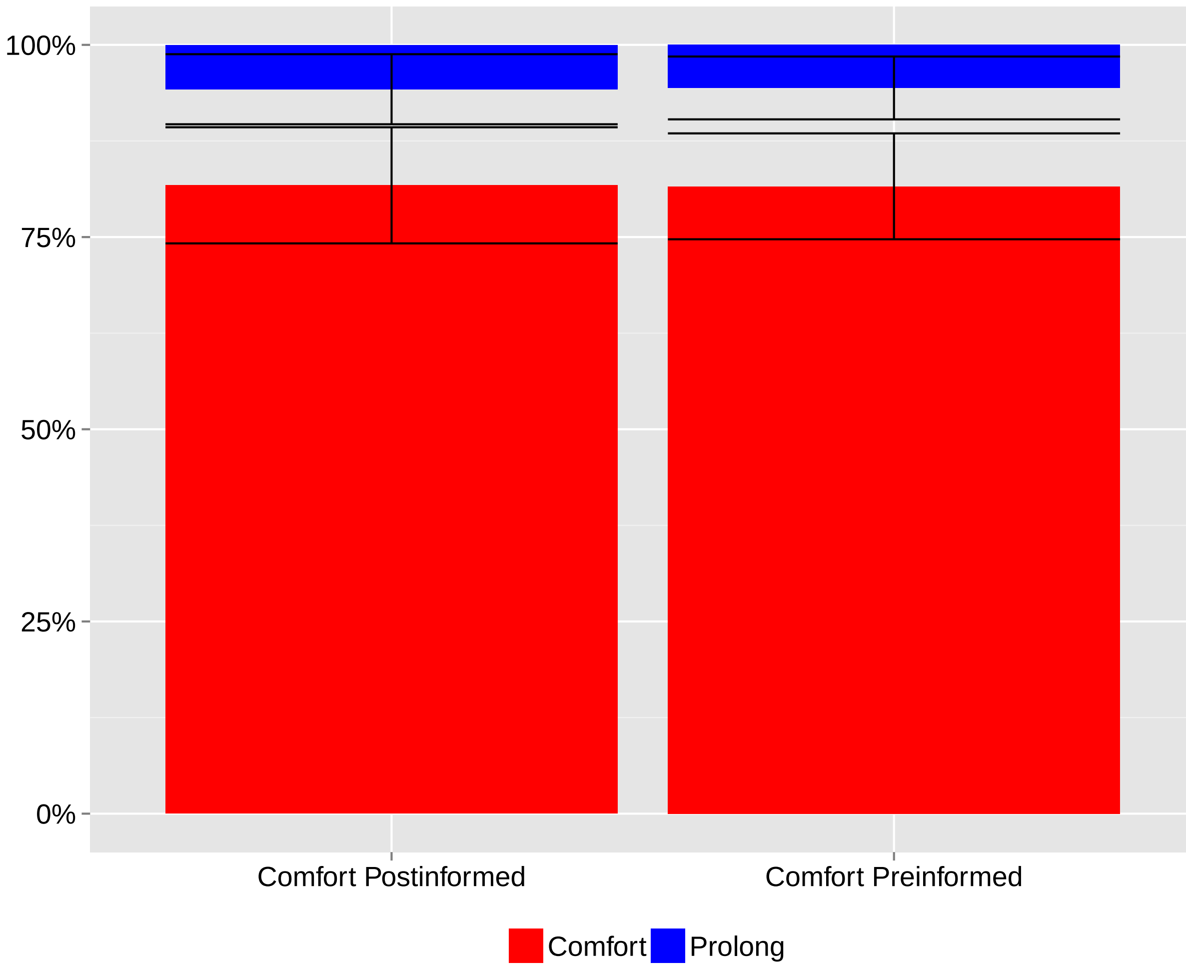
3.4 Capturing Effects from Disclosing Nudges

A detailed description of the results and our analyses of those data are available in the Appendix. Here we summarize our most pertinent findings, which are presented numerically in Table 7 and depicted visually in Figures 8 and 9. Participants showed an overwhelming preference for minimizing discomfort at the end of life rather than prolonging life, especially for the general directives (see Figure 8). When the question was posed in general terms, more than 75% of responses reflected this general goal in all experimental conditions and both choice stages. By comparison, less than 15% of responses selected the goal of prolonging life, with the remaining participants leaving that decision to someone else.

Preferences for comfort in the general directive were so fixed that they were not affected by defaults or disclosure of defaults (that is, choices did not differ by condition in Figure 8). We note that these results differ from recent work using real advance directives (Halpern et al. 2013) in which defaults had a large impact on participants' general goals. One possible explanation is that the highly educated respondents in our study had more definitive preferences about end-of-life care than did the less educated population from the earlier article.

Unlike the results for general directives, defaults for specific treatments, when the participant is only informed after the fact, are effective (see Figure 9A). We could observe this after averaging across the five specific interventions that participants considered: On this combined measure, 46.9% of participants who were given the comfort default (but not informed about it in advance) expressed

Figure 8: The impact of defaults on overall goal for care.



Note. Error bars are included to indicate 95% confidence intervals.

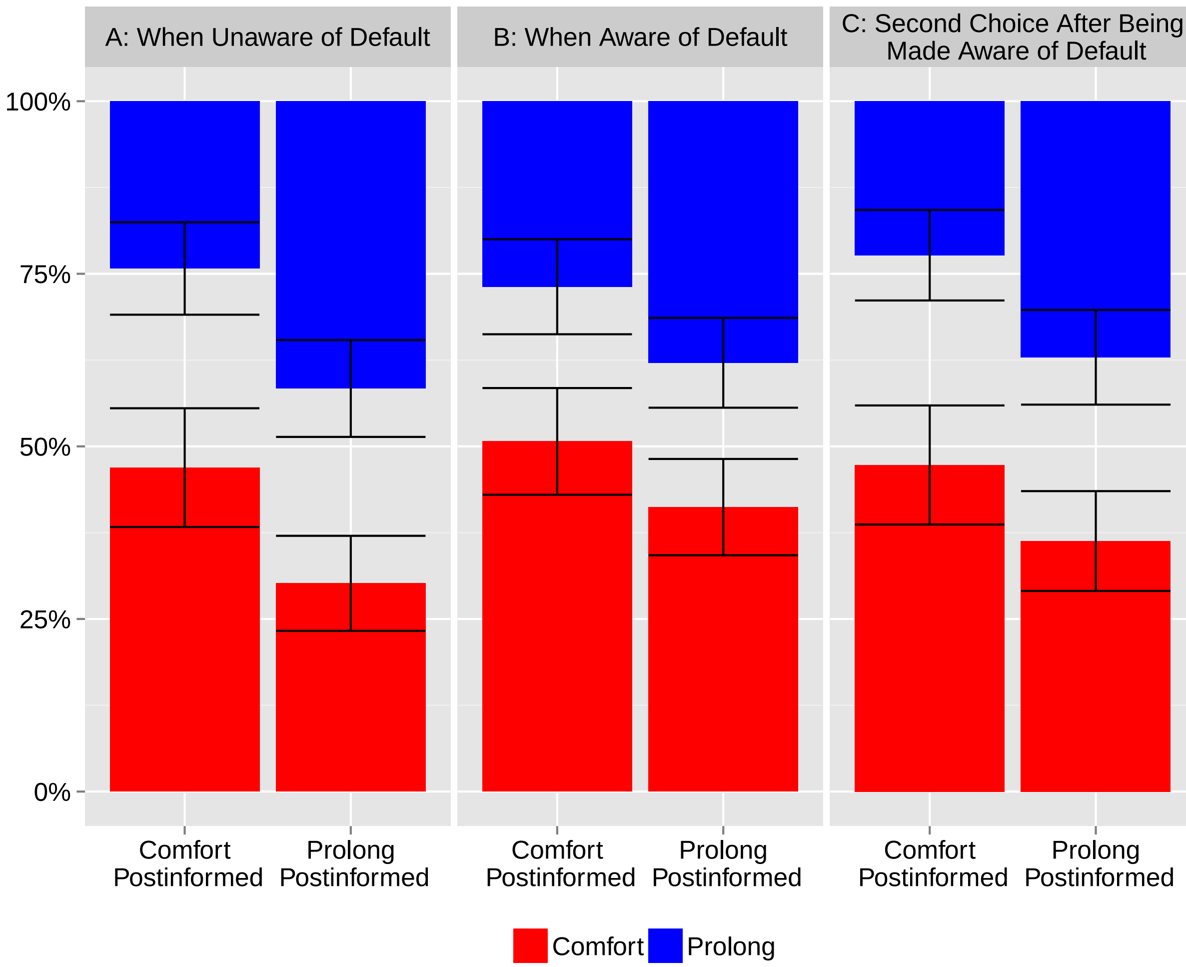
Table 7: Percentage choosing goal and treatment options by stage, default, and condition.

		Choice 1			
		Comfort default		Prolong default	
Question	Choice	Pre-informed	Post-informed	Pre-informed	Post-informed
Overall goal	Choose comfort	81.6%	81.7%	80.5%	78.2%
	Do not choose	12.8%	12.5%	7.5%	16.1%
	Choose prolong	5.6%	5.8%	12.0%	5.6%
Average of 5 specific treatments	Choose comfort	50.7%	46.9%	41.2%	30.2%
	Do not choose	22.4%	28.2%	20.9%	28.2%
	Choose prolong	26.9%	24.2%	37.9%	41.6%
		Choice 2			
		Comfort default		Prolong default	
Question	Choice	Pre-informed	Post-informed	Pre-informed	Post-informed
Overall goal	Choose comfort	76.0%	76.9%	79.7%	79.8%
	Do not choose	12.8%	15.4%	7.5%	14.5%
	Choose prolong	11.2%	7.7%	12.8%	5.6%
Average of 5 specific treatments	Choose comfort	53.8%	47.3%	45.4%	36.3%
	Do not choose	24.6%	30.4%	22.1%	26.6%
	Choose prolong	21.6%	22.3%	32.5%	37.1%

a preference for comfort. By comparison, only 30.2% of those given the prolong default (again with no warning about defaults) expressed a preference for comfort (a difference of 17 percentage points, or 36% [17/46.9]).

The main purpose of the study was to examine the impact on nudge effectiveness of informing people that they were being nudged, a question that is best addressed by analyzing the effects of preinforming people about directive choices. Figure 9B presents the impact of the default when people were preinformed. As can be seen in the figure, preinforming people about defaults weakened but did not wipe out their effectiveness. When participants completed the advance directive after being informed about the impact of the defaults, 50.7% of participants given the comfort default expressed a preference for comfort, compared with only 41.2% of those given the prolong life default (a difference of 10 percentage points, or 19%). Although all specific treatment choices were affected by the default in the predicted direction, the effect is statistically significant only for a single item (dialysis) and for the average of all five items (see the Appendix). Preinforming participants about the default may have weakened its impact, but did not eliminate the default's

Figure 9: The impact of defaults on responses to specific treatments.



Note. Error bars are included to indicate 95% confidence intervals.

effect.

Postinforming people that they have been defaulted and then asking them to choose again in a neutral way, with no further nudge, produces a substantial default effect that is not much smaller than the standard default effect, as seen in Figure 9C. When participants completed the advance directive a second time (this time without a default), having been informed after the fact that they had been defaulted, 47.3% of participants given the comfort default expressed a preference for comfort, compared with only 36.3% of those given the prolong life default (a difference of 11 percentage points, or 23%). Again, postinforming participants about the default and allowing them to change their decision may have weakened its impact, but did not eliminate the default's effect.

These results are important because they suggest that either a preinforming or a postinforming strategy can be effective in both disclosing the presence of a nudge and preserving its effectiveness. In addition, the results provide a conservative estimate of the power of defaults because all respondents who were informed at either stage had, by the second stage, been informed both that they had been randomly selected to be defaulted and that others had been randomly selected to receive alternative defaults. In addition, the second-stage advance directives did not include defaults, so any effect of defaults reflects a carryover effect from the first-stage choice. (More detailed analysis of our results and more information listed by specific treatments are available in the Appendix.)

3.5 Defaults Survive Transparency

Despite extensive research questioning whether advance directives have the intended effect of improving quality of end-of-life care (Connors et al. 1995; Fagerlin and Schneider 2004) they continue to be one of the few and major tools that exist to promote this goal. Combining advance directives with default options could steer people toward the types of comfort options for end-of-life care that many experts recommend and that many people desire for themselves. This study suggests such defaults can be transparently implemented, addressing the concerns of many ethicists without losing defaults' effectiveness.

More broadly, our findings demonstrate that default options are a category of nudges that can have an effect even when people are aware that they are in play. Our results are conservative in two ways. First, not only were respondents informed that they were about to be or had been defaulted, but they also learned that other participants received different defaults, thereby eliminating any implicit recommendation in the default. Given that the nudge continued to have an impact, we can only conjecture that the default effect would have been even more persistent if the warning informed them that they had been defaulted deliberately to the choice that policymakers believe is the best option.

Second, our results are conservative in the sense that the second advance directive that participants completed contained no defaults, so the effect of the initial default had to carry over to the second choice. Our experimental design minimized the added cost of switching: Regardless of whether they wanted to switch, respondents had to provide a second set of responses. Presumably, the impact of the initial default would have been even stronger if switching had required more effort for respondents than sticking with their original response.

What exactly produced the carryover effect remains uncertain. It is possible, and perhaps most interesting, that the prior default led respondents to think about the choice in a different way, specifically in a way that reinforced the rationality of the default they were presented with (consistent with Davidai, Gilovich, and Ross 2012). It is, however, also possible that the respondents were mentally lazy and declined to exert effort to reconsider their previous decisions.

Although the switching costs in our study design were small, such costs may explain why we observed default effects for the specific items but not for the overall goal for care. If respondents were sufficiently concerned about representing their preferences accurately for their overall goal item, they may have been willing to engage in the mental effort to overcome the effect of the default. Finally, it is possible that the carryover from the defaults of stage 1 to the (default-free) responses in stage 2 reflected a desire for consistency (Falk and Zimmermann 2013). If so, then carryover effects would be weaker in real-world contexts involving important decisions. If the practice of informing people that they were being defaulted became widespread, moreover, it is unlikely that either of these default-weakening features would be common. That is because defaults would not be chosen at random and advance directives would be filled out only once, with a disclosed default.

Despite our results, it would be premature to conclude that the impact of nudges will always persist when people are aware of them. Our findings are based on hypothetical advance directives—an appropriate first step in research given both the ethical issues involved and the potential repercussions for choices made regarding preferences for medical care at the end of life. Before embracing the general conclusion that warnings do not eliminate the impact of defaults, further research should examine different types of alerts across different settings. Given how weakly defaults affected overall goals for care in this study, it would especially be fruitful to examine the impact of pre- or postinforming participants in areas in which defaults are observed to have robust impact in the absence of transparency. Those areas include decisionmaking regarding retirement savings and organ donation.

Most generally, our findings suggest that the effectiveness of nudges may not depend on deceiving those who are being nudged. This is good news, because policymakers can satisfy the call for transparency advocated in the House of Lords report (House of Lords 2011) with little diminution in the impact of positive interventions. This could help ease concerns that behavioral interventions are manipulative or involve trickery.

Discussion

Information in economics has long been viewed as a means to an end: useful and desirable to the extent that obtaining it leads to better decisions. In three essays, I have presented empirical evidence that this view fails to capture our complex relationship with information. Rather than merely informing our decisions, learning something about ourselves can be painful and aversive, leading people to prefer ignorance even when knowing might lead them to make better decisions (Chapter

1). Informing people about nudges, which are ideally complementary to, but are not intended to be substitutes for, economic incentives, can undermine support for the latter (Chapter 2). However, learning that one has been assigned to a randomly selected default option in a (hypothetical) advance directive does not diminish the impact of the default (Chapter 3).

The first chapter showed that the desire to avoid potentially useful information, previously found in laboratory and field experiments, is a stable and distinct personality trait that can be measured. Far from being limited to a small number of people, I found that the majority of people in my studies desire to avoid information in some instances. Information avoidance seems to be a general trait, although there does seem to be a domain-specific component; for example, some respondents were receptive to information about their health, but did not want to learn about their finances, while others were exhibited the reverse pattern. The chapter presented evidence that this desire to obtain (or avoid) information is stable over time and is distinct from measures such as curiosity or the need for cognition. Importantly for economists, the desire to obtain information is positively correlated with a willingness to take risks, and with patience. Learning something that could be either favorable or unfavorable may be like a gamble that has some probability of returning a gain and some chance of a loss. And while the costs of learning something painful are immediate, the benefits from making more informed decisions materialize in the future.

In the second chapter, I reported the first evidence of policy crowding-out, in which introducing a nudge undermines support for a more effective carbon tax. In addition to showing crowding-out, all of the studies also found that participants, including policymakers, substantially overestimated the effect of nudges, as compared with more substantive policies. These observations applied even to people directly involved with policy-making, who were no less susceptible to the crowding-out effect, and who also overestimated the impact of a nudge relative to a more substantive policy. In the final study of the series, I found, however, that correcting people's beliefs about the nudge's effectiveness prevented it from crowding-out support for the carbon tax, without harming support for the nudge.

The third chapter tackled a concern among policymakers that information about nudges may render them ineffective. I found that assigning participants to defaults in a hypothetical advance directive influences the choices they make. The effect of the default, however, was not diminished by disclosing that participants were randomly assigned to one of the options. Moreover, when people learned that they had been defaulted and were given a chance to fill out the advance directive again (without a default), their choices remained unchanged. That is, providing information either before or after nudging people with a default did not undermine the effectiveness of the intervention and did not create backlash.

In 1961, George Stigler ended his seminal paper observing that "our understanding of economic life will be incomplete if we do not systematically take account of the cold winds of ignorance." The work in this dissertation found that the methods of psychology can be productively used by economists to bring warmth into our understanding of the economics of information.

Bibliography

Acquisti, Alessandro, Leslie K. John, and George Loewenstein. 2013. "What Is Privacy Worth?" *The Journal of Legal Studies* 42 (2): 249–74. https://doi.org/10.1086/671754.

Ajzen, Icek, and Martin Fishbein. 1977. "Attitude-Behavior Relations: A Theoretical Analysis and Review of Empirical Research." *Psychological Bulletin* 84 (5): 888–918. https://doi.org/10.1037/0033-2909.84.5.888.

Allcott, Hunt. 2011. "Social Norms and Energy Conservation." *Journal of Public Economics*, Special issue: The role of firms in tax systems, 95 (9): 1082–95. https://doi.org/10.1016/j.jpubeco.2011.03.003.

———. 2015. "Site Selection Bias in Program Evaluation." *The Quarterly Journal of Economics*

130 (3): 1117–65. https://doi.org/10.1093/qje/qjv015.

Allcott, Hunt, and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *The American Economic Review* 104 (10): 3003–37. https://doi.org/10.1257/aer.104.10.3003.

Antonovsky, Anton. 1993. "The Structure and Properties of the Sense of Coherence Scale." *Social Science & Medicine* 36 (6): 725–33. https://doi.org/10.1016/0277-9536(93)90033-Z.

Asparouhov, Tihomir, and Bengt Muthén. 2009. "Exploratory Structural Equation Modeling." *Structural Equation Modeling: A Multidisciplinary Journal* 16 (3): 397–438. https://doi.org/10.1080/10705510903008204.

Azen, Razia, and David V. Budescu. 2003. "The Dominance Analysis Approach for Comparing Predictors in Multiple Regression." *Psychological Methods* 8 (2): 129–48. https://doi.org/10.1037/1082-989X.8.2.129.

Azen, R., and N. Traxel. 2009. "Using Dominance Analysis to Determine Predictor Importance in Logistic Regression." *Journal of Educational and Behavioral Statistics* 34 (3): 319–47. https://doi.org/10.3102/1076998609332754.

Bandura, Albert. 1986. "Fearful Expectations and Avoidance Actions and Coeffects of Perceived Self-Inefficacy." *American Psychologist* 41: 1389–91.

Bazerman, Max H., and Ovul Sezer. 2016. "Bounded Awareness: Implications for Ethical Decision Making." *Organizational Behavior and Human Decision Processes*, Celebrating fifty years of organizational behavior and decision making research (1966-2016), 136: 95–105. https://doi.org/10.1016/j.obhdp.2015.11.004.

Benartzi, Shlomo, John Beshears, Katherine L. Milkman, Cass R. Sunstein, Richard H. Thaler, Maya Shankar, Will Tucker-Ray, William J. Congdon, and Steven Galing. 2017. "Should Governments Invest More in Nudging?" *Psychological Science* 28 (8): 1041–55. https://doi.org/10.1177/0956797617702501.

Bergman, Peter, and Todd Rogers. 2017. "The Impact of Defaults on Technology Adoption, and Its Underappreciation by Policymakers." Faculty Research Working Paper Series. Harvard Kennedy School.

Bovens, Luc. 2009. "The Ethics of Nudge." In *Preference Change: Approaches from Philosophy, Economics and Psychology*, edited by Till Grüne-Yanoff and Sven Ove Hansson, 207–19. Springer.

Cacioppo, John T, Richard E Petty, and Chuan Feng Kao. 1984. "The Efficient Assessment of Need for Cognition." *Journal of Personality Assessment* 48: 306–307.

Camerer, Colin, Samuel Issacharoff, George Loewenstein, Ted O'Donoghue, and Matthew Rabin. 2003a. "Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism"." *University of Pennsylvania Law Review* 151 (3): 1211–54. https://doi.org/10.2307/3312889.

———. 2003b. "Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism"." *University of Pennsylvania Law Review* 151 (3): 1211–54. https://doi.org/10.2307/3312889.

Caplin, Andrew, and Kfir Eliaz. 2003. "AIDS Policy and Psychology: A Mechanism-Design Approach." *RAND Journal of Economics* 34 (4): 631–46.

Caprara, Gian Vittorio, Michele Vecchione, Guido Alessandri, Maria Gerbino, and Claudio Barbaranelli. 2011. "The Contribution of Personality Traits and Self-Efficacy Beliefs to Academic Achievement: A Longitudinal Study: Personality Traits, Self-Efficacy Beliefs and Academic Achievement." *British Journal of Educational Psychology* 81 (1): 78–96. https://doi.org/10.1348/2044-8279.002004.

Cialdini, Robert B., Melanie R. Trost, and Jason T. Newsom. 1995. "Preference for Consistency: The Development of a Valid Measure and the Discovery of Surprising Behavioral Implications." *Journal of Personality and Social Psychology* 69 (2): 318.

Connors, Alfred F., Neal V. Dawson, Norman A. Desbiens, William J. Fulkerson, Lee Goldman, William A. Knaus, Joanne Lynn, et al. 1995. "A Controlled Trial to Improve Care for Seriously Ill Hospitalized Patients: The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT)." *JAMA* 274 (20): 1591–8. https://doi.org/10.1001/jama.1995.03530200027032.

Cox, Louis A. 1985. "A New Measure of Attributable Risk for Public Health Applications." *Management Science* 31 (7): 800–813. https://doi.org/10.1287/mnsc.31.7.800.

Davidai, S., T. Gilovich, and L. D. Ross. 2012. "The Meaning of Default Options for Potential Organ Donors." *Proceedings of the National Academy of Sciences* 109 (38): 15201–5. https://doi.org/10.1073/pnas.1211695109.

Davidai, Shai, and Eldar Shafir. 2018. "Are 'Nudges' Getting a Fair Shot? Joint Versus Separate Evaluation." *Behavioural Public Policy*, April, 1–19. https://doi.org/10.1017/bpp.2018.9.

Dolan, Paul, and Matteo M. Galizzi. 2015. "Like Ripples on a Pond: Behavioral Spillovers and Their Implications for Research and Policy." *Journal of Economic Psychology* 47 (April): 1–16. https://doi.org/10.1016/j.joep.2014.12.003.

Druckman, James N., Erik Peterson, and Rune Slothuus. 2013. "How Elite Partisan Polarization Affects Public Opinion Formation." *American Political Science Review* 107 (01): 57–79. https://doi.org/10.1017/S0003055412000500.

Eil, David, and Justin M Rao. 2011. "The Good News-Bad News Effect: Asymmetric Processing of Objective Information About Yourself." *American Economic Journal: Microeconomics* 3 (2): 114–38. https://doi.org/10.1257/mic.3.2.114.

Elbel, Brian, Rogan Kersh, Victoria L. Brescoll, and L. Beth Dixon. 2009. "Calorie Labeling and Food Choices: A First Look at the Effects on Low-Income People in New York City." *Health Affairs* 28 (6): w1110–w1121. https://doi.org/10.1377/hlthaff.28.6.w1110.

Engelhardt, Gary V, and Jonathan Gruber. 2004. "Social Security and the Evolution of Elderly

Poverty." Working Paper 10466. National Bureau of Economic Research. https://doi.org/10.3386/w10466.

Estrella, Arturo. 1998. "A New Measure of Fit for Equations with Dichotomous Dependent Variables." *Journal of Business & Economic Statistics* 16 (2): 198–205. https://doi.org/10.1080/07350015.1998.10524753.

Fagerlin, Angela, and Carl E. Schneider. 2004. "Enough: The Failure of the Living Will." *Hastings Center Report* 34 (2): 30–42. https://doi.org/10.2307/3527683.

Falk, Armin, and Florian Zimmermann. 2013. "A Taste for Consistency and Survey Response Behavior." *CESifo Economic Studies* 59 (1): 181–93. https://doi.org/10.1093/cesifo/ifs039.

———. 2014. "Beliefs and Utility: Experimental Evidence on Preferences for Information." Working paper.

Fernandes, Daniel, John G. Lynch, and Richard G. Netemeyer. 2014. "Financial Literacy, Financial Education, and Downstream Financial Behaviors." *Management Science* 60 (8): 1861–83. https://doi.org/10.1287/mnsc.2013.1849.

Festinger, Leon. 1962. "Cognitive Dissonance." *Scientific American* 207 (4): 93–106.

Fishbach, Ayelet, Ravi Dhar, and Ying Zhang. 2006. "Subgoals as Substitutes or Complements: The Role of Goal Accessibility." *Journal of Personality and Social Psychology* 91 (2): 232–42. https://doi.org/10.1037/0022-3514.91.2.232.

Flora, David B., and Patrick J. Curran. 2004. "An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis with Ordinal Data." *Psychological Methods* 9 (4): 466–91. https://doi.org/10.1037/1082-989X.9.4.466.

Ganguly, Ananda, and Joshua Tasoff. 2016. "Fantasy and Dread: The Demand for Information and the Consumption Utility of the Future." *Management Science*, October. https://doi.org/10.1287/mnsc.2016.2550.

Ginsburg, Geoffrey S., and Jeanette J. McCarthy. 2001. "Personalized Medicine: Revolutionizing Drug Discovery and Patient Care." *TRENDS in Biotechnology* 19 (12): 491–96.

Gneezy, Uri, and Jan Potters. 1997. "An Experiment on Risk Taking and Evaluation Periods." *The Quarterly Journal of Economics* 112 (2): 631–45.

Golman, Russell, David Hagmann, and George Loewenstein. 2017. "Information Avoidance." *Journal of Economic Literature* 55 (1): 96–135. https://doi.org/10.1257/jel.20151245.

Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu. n.d. "How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice." *Political Analysis*.

Haisley, Emily, Kevin G. Volpp, Thomas Pellathy, and George Loewenstein. n.d. "The Impact of Alternative Incentive Schemes on Completion of Health Risk Assessments." *American Journal of Health Promotion: AJHP* 26 (3): 184–88. https://doi.org/10.4278/ajhp.100729-ARB-257.

Hall, Marissa G., Theresa M. Marteau, Cass R. Sunstein, Kurt M. Ribisl, Seth M. Noar, Elizabeth N. Orlan, and Noel T. Brewer. 2018. "Public Support for Pictorial Warnings on Cigarette Packs: An Experimental Study of US Smokers." *Journal of Behavioral Medicine*, February, 1–8. https://doi.org/10.1007/s10865-018-9910-2.

Halpern, Scott D., Peter A. Ubel, and David A. Asch. 2007. "Harnessing the Power of Default Options to Improve Health Care." *The New England Journal of Medicine* 357 (13): 1340–4. https://doi.org/http://dx.doi.org/10.1056/NEJMsb071595.

Halpern, S. D., G. Loewenstein, K. G. Volpp, E. Cooney, K. Vranas, C. M. Quill, M. S. McKenzie, et al. 2013. "Default Options in Advance Directives Influence How Patients Set Goals for End-of-Life Care." *Health Affairs* 32 (2): 408–17. https://doi.org/10.1377/hlthaff.2012.0895.

Hansen, J., S. Marx, and Elke Weber. 2004. "The Role of Climate Change Perceptions, Expectations, and Forecasts in Farmer Decision Making: The Argentine Pampas and South Florida." Palisades, New York.: International Research Institute for Climate Prediction.

Hedlin, Simon, and Cass Sunstein. 2016. "Does Active Choosing Promote Green Energy Use? Experimental Evidence." https://doi.org/10.15779/z387g30.

Hess, Stephane, Mark Fowler, Thomas Adler, and Aniss Bahreinian. 2012. "A Joint Model for Vehicle Type and Fuel Type Choice: Evidence from a Cross-Nested Logit Study." *Transportation* 39 (3): 593–625. https://doi.org/10.1007/s11116-011-9366-5.

Ho, Emily, David Budescu, and Han-Hui Por. 2017. "Psychological Challenges in Communicating About Climate Change and Its Uncertainties." *Oxford Research Encyclopedia of Climate Change*.

Hooper, Daire, Joseph Coughlan, and Michael Mullen. 2008. "Structural Equation Modelling: Guidelines for Determining Model Fit." *Articles*, 2.

House of Lords. 2011. "Science and Technology Select Committee Report." *Journal of Radiological Protection : Official Journal of the Society for Radiological Protection* 26 (July): 331. https://doi.org/10.1088/0952-4746/26/3/L01.

Howell, Jennifer L., and James A. Shepperd. 2016. "Establishing an Information Avoidance Scale." *Psychological Assessment* 28 (12): 1695–1708. https://doi.org/10.1037/pas0000315.

Hu, Li?tze, and Peter M. Bentler. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives." *Structural Equation Modeling: A Multidisciplinary Journal* 6 (1): 1–55. https://doi.org/10.1080/10705519909540118.

Huckfeldt, Robert, and John Sprague. 1987. "Networks in Context: The Social Flow of Political Information." *The American Political Science Review* 81 (4): 1197–1216.

Interagency Working Group, and others. 2013. "Technical Update on the Social Cost of Carbon for Regulatory Impact Analysis-Under Executive Order 12866." Washington, D.C.: Office of Management and Budget.

Jethwani, Kamal, Julie Kvedar, and Joseph Kvedar. 2010. "Behavioral Phenotyping: A Tool for

Personalized Medicine." *Personalized Medicine* 7 (6): 689–93. https://doi.org/10.2217/pme.10.62.

John, Oliver P., and Sanjay Srivastava. 1999. "The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives." *Handbook of Personality: Theory and Research* 2 (1999): 102–38.

Johnson, Eric J., and Daniel Goldstein. 2003. "Do Defaults Save Lives?" *Science*, 1338–9.

Johnson, Eric J., and Daniel G. Goldstein. 2004. "Defaults and Donation Decisions:" *Transplantation* 78 (12): 1713–6. https://doi.org/10.1097/01.TP.0000149788.10382.B2.

Johnson, Eric J., John Hershey, Jacqueline Meszaros, and Howard Kunreuther. 1993. "Framing, Probability Distortions, and Insurance Decisions." *Journal of Risk and Uncertainty* 7 (1): 35–51. https://doi.org/10.1007/BF01065313.

Kahan, Dan M., Asheley Landrum, Katie Carpenter, Laura Helft, and Kathleen Hall Jamieson. 2017. "Science Curiosity and Political Information Processing: Curiosity and Information Processing." *Political Psychology* 38 (February): 179–99. https://doi.org/10.1111/pops.12396.

Kaiser, Henry F. 1960. "The Application of Electronic Computers to Factor Analysis." *Educational and Psychological Measurement* 20 (1): 141–51. https://doi.org/10.1177/001316446002000116.

Kashdan, Todd B., Matthew W. Gallagher, Paul J. Silvia, Beate P. Winterstein, William E. Breen, Daniel Terhar, and Michael F. Steger. 2009. "The Curiosity and Exploration Inventory-II: Development, Factor Structure, and Psychometrics." *Journal of Research in Personality* 43 (6): 987–98. https://doi.org/10.1016/j.jrp.2009.04.011.

Kirby, Kris N., Nancy M. Petry, and Warren K. Bickel. 1999. "Heroin Addicts Have Higher Discount Rates for Delayed Rewards Than Non-Drug-Using Controls." *Journal of Experimental Psychology: General* 128 (1): 78–87. https://doi.org/10.1037/0096-3445.128.1.78.

Koszegi, Botond, and Matthew Rabin. 2006. "A Model of Reference-Dependent Preferences." *The Quarterly Journal of Economics* 121 (4): 1133–65.

Kressel, Laura M., and Gretchen B. Chapman. 2007. "The Default Effect in End-of-Life Medical Treatment Preferences." *Medical Decision Making* 27 (3): 299–310. https://doi.org/10.1177/0272989X07300608.

Kressel, Laura M., Gretchen B. Chapman, and Elaine Leventhal. 2007. "The Influence of Default Options on the Expression of End-of-Life Treatment Preferences in Advance Directives." *Journal of General Internal Medicine* 22 (7): 1007–10. https://doi.org/10.1007/s11606-007-0204-6.

Kruglanski, Arie W. 2013. *Lay Epistemics and Human Knowledge: Cognitive and Motivational Bases.* Place of publication not identified: Springer-Verlag New York.

Lee, Sunhee, and Howard J. Klein. 2002. "Relationships Between Conscientiousness, Self-Efficacy, Self-Deception, and Learning over Time." *Journal of Applied Psychology* 87 (6): 1175–82. https://doi.org/10.1037/0021-9010.87.6.1175.

Levitt, Steven D., John A. List, Susanne Neckermann, and Sally Sadoff. 2016. "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance." *American Economic Journal: Economic Policy* 8 (4): 183–219. https://doi.org/10.1257/pol.20130358.

Loewenstein, George. 1994. "The Psychology of Curiosity: A Review and Reinterpretation." *Psychological Bulletin* 116 (1): 75–98.

———. 2006. "The Pleasures and Pains of Information." *Science* 312 (5774): 704–6. https://doi.org/10.1126/science.1128388.

Loewenstein, George, and Nick Chater. 2017a. "Putting Nudges in Perspective." *Behavioural Public Policy* 1 (1): 26–53. https://doi.org/10.1017/bpp.2016.7.

———. 2017b. “Putting Nudges in Perspective.” *Behavioural Public Policy* 1 (01): 26–53. https://doi.org/10.1017/bpp.2016.7.

Loewenstein, George, and Peter Ubel. 2010. “Economics Behaving Badly.” *The New York Times*, July.

Madrian, Brigitte C., Hal E. Hershfield, Abigail B. Sussman, Saurabh Bhargava, Jeremy Burke, Scott A. Huettel, Julian Jamison, et al. 2017. “Behaviorally Informed Policies for Household Financial Decisionmaking.” *Behavioral Science & Policy* 3 (1): 26–40. https://doi.org/10.1353/bsp.2017.0003.

Madrian, Brigitte C., and Dennis F. Shea. 2001. “The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior.” *The Quarterly Journal of Economics* 116 (4): 1149–87. https://doi.org/10.1162/003355301753265543.

Mardia, K. V. 1970. “Measures of Multivariate Skewness and Kurtosis with Applications.” *Biometrika* 57 (3): 519–30. https://doi.org/10.2307/2334770.

Marshall, George. 2014. *Don’t Even Think About It: Why Our Brains Are Wired to Ignore Climate Change*. 1 edition. Bloomsbury USA.

Marteau, T. M., D. Ogilvie, M. Roland, M. Suhrcke, and M. P. Kelly. 2011. “Judging Nudging: Can Nudging Improve Population Health?” *BMJ* 342 (jan25 3): d228–d228. https://doi.org/10.1136/bmj.d228.

Martocchio, Joseph J., and Timothy A. Judge. 1997. “Relationship Between Conscientiousness and Learning in Employee Training: Mediating Influences of Self-Deception and Self-Efficacy.” *Journal of Applied Psychology* 82 (5): 764.

Maul, Andrew. 2017. “Rethinking Traditional Methods of Survey Validation.” *Measurement: Interdisciplinary Research and Perspectives*, August, 1–19. https://doi.org/10.1080/15366367.2017.1348108.

McFadden, Daniel. 1974. "The Measurement of Urban Travel Demand." *Journal of Public Economics* 3 (4): 303–28. https://doi.org/10.1016/0047-2727(74)90003-6.

McKenzie, C. R.M., M. J. Liersch, and S. R. Finkelstein. 2006. "Recommendations Implicit in Policy Defaults." *Psychological Science* 17 (5): 414–20. https://doi.org/10.1111/j.1467-9280.2006.01721.x.

Meijers, Marijn H. C., Peeter W. J. Verlegh, Marret K. Noordewier, and Edith G. Smit. 2015. "The Dark Side of Donating: How Donating May License Environmentally Unfriendly Behavior." *Social Influence* 10 (4): 250–63. https://doi.org/10.1080/15534510.2015.1092468.

Minson, Julia A, Catherine H. Tinsley, and Frances S. Chen. 2018. "Why Won't You Listen to Me? Measuring Receptiveness to Opposing Views." Working Paper. Harvard Kennedy School.

Mischel, Waltę, and Yuichi Shoda. 1995. "A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics." *Psychological Review (PsycARTICLES)* 102: 2.

Nagelkerke, N. J. D. 1991. "A Note on a General Definition of the Coefficient of Determination." *Biometrika* 78 (3): 691–92. https://doi.org/10.1093/biomet/78.3.691.

Nash, Nick, Lorraine Whitmarsh, Stuart Capstick, Tom Hargreaves, Wouter Poortinga, Gregory Thomas, Elena Sautkina, and Dimitrios Xenias. 2017. "Climate-Relevant Behavioral Spillover and the Potential Contribution of Social Practice Theory." *Wiley Interdisciplinary Reviews: Climate Change* 8 (6): e481. https://doi.org/10.1002/wcc.481.

Oster, Emily, Ira Shoulson, and E. Ray Dorsey. 2013. "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease." *The American Economic Review* 103 (2): 804–30.

Patel, Mitesh S., Susan C. Day, Scott D. Halpern, C. William Hanson, Joseph R. Martinez, Steven Honeywell, and Kevin G. Volpp. 2016. "Generic Medication Prescription Rates After Health SystemWide Redesign of Default Options Within the Electronic Health Record." *JAMA*

Internal Medicine 176 (6): 847–48. https://doi.org/10.1001/jamainternmed.2016.1691.

"Policymakers Around the World Are Embracing Behavioural Science." 2017. *The Economist*, May.

Reisch, Lucia A., and Cass R. Sunstein. 2016. "Do Europeans Like Nudges?" *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2739118.

Rogers, Todd, and Avi Feller. n.d. "Reducing Student Absences at Scale by Targeting Parents' Misbeliefs." *Nature Human Behaviour*, 12.

Schork, Nicholas. 2015. "Time for One-Person Trials." *NATURE* 520.

Schunk, Dale H. 1990. "Goal Setting and Self-Efficacy During Self-Regulated Learning." *Educational Psychologist* 25 (1): 71–86.

Schwarzer, R, and M Jerusalem. 1995. "Generalized Self-Efficacy Scale." In *Measures in Health Psychology: A User's Portfolio. Causal and Control Beliefs.*, edited by & M. Johnston J. Weinman S. Wright, 35–37. Windsor, UK: NFER NELSON.

Sicherman, Nachum, George Loewenstein, Duane J. Seppi, and Stephen P. Utkus. 2016. "Financial Attention." *The Review of Financial Studies* 29 (4): 863–97. https://doi.org/10.1093/rfs/hhv073.

Steiger, J. H., and J. C. Lind. 1980. "Statistically Based Tests for the Number of Factors." In *Psychometric Society*. Iowa City, IA.

Stigler, George J. 1961. "The Economics of Information." *Journal of Political Economy* 69 (3): 213–25.

Sullivan, Patrick S., Amy Lansky, Amy Drake, and HITS-2000 Investigators. 2004. "Failure to Return for HIV Test Results Among Persons at High Risk for HIV Infection: Results from a Multistate Interview Project." *Journal of Acquired Immune Deficiency Syndromes (1999)* 35 (5): 511–18.

Sunstein, Cass R. 2016. "Do People Like Nudges?" *SSRN*, February.

———. 2017. "Do People Like Nudges?" SSRN Scholarly Paper ID 2604084. Rochester, NY: Social Science Research Network.

Swan, Melanie. 2009. "Emerging Patient-Driven Health Care Models: An Examination of Health Social Networks, Consumer Personalized Medicine and Quantified Self-Tracking." *International Journal of Environmental Research and Public Health* 6 (2): 492–525. https://doi.org/10.3390/ijerph6020492.

Sweeny, Kate, Darya Melnyk, Wendi Miller, and James A. Shepperd. 2010. "Information Avoidance: Who, What, When, and Why." *Review of General Psychology* 14 (4): 340–53. https://doi.org/10.1037/a0021288.

Thaler, Richard H., and Shlomo Benartzi. 2004. "Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving." *Journal of Political Economy* 112 (S1): S164–S187. https://doi.org/10.1086/380085.

Thaler, Richard H., and Cass R. Sunstein. 2003. "Libertarian Paternalism." *The American Economic Review* 93 (2): 175–79.

———. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press.

Thøgersen, John. 1999. "Spillover Processes in the Development of a Sustainable Consumption Pattern." *Journal of Economic Psychology* 20 (1): 53–81. https://doi.org/10.1016/S0167-4870(98)00043-9.

Tiefenbeck, Verena, Thorsten Staake, Kurt Roth, and Olga Sachs. 2013. "For Better or for Worse? Empirical Evidence of Moral Licensing in a Behavioral Energy Conservation Campaign." *Energy Policy* 57 (June): 160–71. https://doi.org/10.1016/j.enpol.2013.01.021.

Truelove, Heather Barnes, Amanda R. Carrico, Elke U. Weber, Kaitlin Toner Raimi, and Michael

P. Vandenbergh. 2014. “Positive and Negative Spillover of Pro-Environmental Behavior: An Integrative Review and Theoretical Framework.” *Global Environmental Change* 29 (November): 127–38. https://doi.org/10.1016/j.gloenvcha.2014.09.004.

Tucker, Ledyard R, and Charles Lewis. 1973. “A Reliability Coefficient for Maximum Likelihood Factor Analysis.” *Psychometrika* 38 (1): 1–10. https://doi.org/10.1007/BF02291170.

Weber, Elke. 1997. “Perception and Epectation of Climate Change: Precondition for Economic and Technological Adaptation.” In *Psychological Perspectives to Environmental and Ethical Issues in Management*, edited by Max Bazerman, D. Messick, A. Tensbrunsel, and K. Wade-Benzoni, 314–41. Jossey-Bass, San Francisco.

Webster, Donna M., and Arie W. Kruglanski. 1994. “Individual Differences in Need for Cognitive Closure.” *Journal of Personality and Social Psychology* 67 (6): 1049–62. https://doi.org/10.1037/0022-3514.67.6.1049.

Werfel, Seth H. 2017. “Household Behaviour Crowds Out Support for Climate Change Policy When Sufficient Progress Is Perceived.” *Nature Climate Change* 7 (7): 512–15. https://doi.org/10.1038/nclimate3316.

Wortman, Camille B., and Jack W. Brehm. 1975. “Responses to Uncontrollable Outcomes: An Integration of Reactance Theory and the Learned Helplessness Model.” Edited by Leonard Berkowitz. *Advances in Experimental Social Psychology* Volume 8: 277–336.

Yoeli, Erez, David V. Budescu, Amanda R. Carrico, Magali A. Delmas, J. R. DeShazo, Paul J. Ferraro, Hale A. Forster, et al. 2017. “Behavioral Science Tools to Strengthen Energy & Environmental Policy.” *Behavioral Science & Policy* 3 (1): 68–79. https://doi.org/10.1353/bsp.2017.0006.

YouGov. 2017. “YouGov | Most Americans Support DACA - but Trump’s Base Opposes.” *YouGov: What the World Thinks*. //today.yougov.com/news/2017/09/05/most-americans-supports-daca-trumps-base-opposes/.

Zeelenberg, Marcel. 1999. "Anticipated Regret, Expected Feedback and Behavioral Decision Making." *Journal of Behavioral Decision Making* 12 (2): 93.

A Appendix: The Hidden Cost of Soft Paternalism

A.1 Study 1A

When participants could only implement the tax, we find that 70.3% supported its implementation. When an opportunity to implement the nudge was also available, 45.00% supported implementing both, 32.00% chose to implement the nudge only, 10.00% implemented only the tax, and 13.00% desired neither option. This implies that only 55% supported implementing the tax or both the tax and the nudge, which is significantly less support than in the policy only condition ($\chi^2(1, n = 201) = 4.39$, $p = .036$).

In Table A.1, we present a series of logistic regressions in which the dependent variable is 1 if the respondent decided to implement either the tax only or both the tax and the nudge, and 0 otherwise. Ideally, we could rely on cross-nested logistic regression to account for the fact that choosing to implement "both" policies in the condition that allows implementing the tax and the nudge is a stronger policy response than merely implementing the tax when both options are available. Moreover, for participants in the "Tax Only" condition, choosing to implement the tax does not tell us whether they would also have wanted to implement the nudge. However, we are not sufficiently powered to achieve model convergence for these more flexible models, which are ordinarily employed when the number of observations exceeds 10,000.(Hess et al. 2012)

In our baseline Model 1, we show that the results from our chi-square analysis also hold in the regression framework. Introducing the option of a nudge decreases the log-odds of supporting the tax by half ($p < 0.05$). In Model 2, we add control variables for how effective and how painful participants thought the tax to be. Those who thought the tax was more effective were, not

Table A.1: Logistic regression on the decision to implement the tax (or both the tax and the nudge). Introducing an option to implement a nudge decreases the likelihood of implementing the tax. This holds with a series of controls, including how effective and painful they rate the tax policy. Notably, effectiveness or painfulness of the nudge do not influence support for the tax.

	Model 1	Model 2	Model 3	Model 4
Nudge Available	-0.661^{*}	-1.211^{***}	-1.207^{***}	-1.428^{***}
	(0.296)	(0.366)	(0.366)	(0.401)
Tax Effectiveness		0.791^{***}	0.775^{***}	0.711^{***}
		(0.166)	(0.187)	(0.202)
Tax Painfulness		-0.658^{***}	-0.647^{***}	-0.731^{***}
		(0.175)	(0.180)	(0.194)
Nudge Effectiveness			0.024	-0.091
			(0.199)	(0.214)
Nudge Painfulness			-0.067	0.245
			(0.214)	(0.241)
Climate Change Exists				0.246
				(0.242)
Conservative				-1.673^{***}
				(0.409)
(Intercept)	0.861^{***}	0.497	0.548	0.621
	(0.218)	(0.746)	(0.882)	(1.405)
Log Likelihood	-130.255	-102.000	-101.940	-90.632
Num. obs.	201	201	201	201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Coefficients are expected log-odds.

surprisingly, more likely to support its implementation, while those who believed it to be more painful were less likely to do so. The main effect of introducing the nudge remains significant ($p < 0.001$), now suggesting the nudge reduced the log-odds of supporting the tax by 65%. Models 3 and 4 add further controls and show that the main effect is not diminished by doing so. Notably, we find that how effective or painful participants rated the nudge to be did not affect their decision to implement the tax.

A.2 Study 1B

We first conduct a manipulation check to see whether highlighting the (obvious) costs of a carbon tax to consumers lead participants to rate the tax as more painful. Indeed, the painfulness rating (on a five-point Likert scale) increases by a full point, from 2.06 to 3.16 ($t(797.79) = -13.88$, $p < .001$). The framing manipulation, importantly, does not affect the perceived effectiveness of the tax in Low Pain (3.12) and High Pain conditions (3.00, $t(797.54) = 1.51$, $p = .132$). Moreover, the pain framing is overall effective at decreasing support for the tax, from 67.58% to 41.6% ($t(795.45) = 7.63$, $p < .001$).

Table A.2 presents results from a series of logistic regressions on the decision to implement the tax. We observe significant main effects of our two conditions, painfulness and choice set (Model 1) when controlling for perceived effectiveness of the tax. Reading about a painful tax policy reduces support for the tax, as does introducing a nudge into the choice set. In Model 1, introducing the tax reduces support by 36%, while the painful framing reduces support by 70%. The interaction between the two conditions is non-significant (Model 2), implying the crowding out effect observed cannot be explained by the perceived painfulness of the tax. The main effect of both painfulness and choice set remain significant when controlling for perceived policy painfulness and effectiveness, political affiliation and the interaction of our experimental conditions (Models 3 and 4). Our best-fitting Model 4 estimates that the decrease of support for the tax from introducing a nudge (42%) is approximately the same as from the painful framing (46%).

Table A.2: Logistic regression for Study 1B. We observe that framing the tax as more painful decreases willingness to implement it. Notably, we see no interaction effect between our choice set manipulation and the high pain framing. Model 4 omits the interaction term for a less complex model that provides equivalent fit.

	Model 1	Model 2	Model 3	Model 4
Tax + Nudge	-0.354^{*}	-0.390	-0.495^{**}	-0.546^{**}
	(0.148)	(0.215)	(0.175)	(0.179)
High Pain	-1.084^{***}	-1.118^{***}	-0.533^{**}	-0.623^{**}
	(0.148)	(0.212)	(0.189)	(0.195)
Tax + Nudge x High Pain		0.067		
		(0.296)		
Tax Effectiveness			0.820^{***}	0.802^{***}
			(0.094)	(0.096)
Tax Painfulness			-0.757^{***}	-0.758^{***}
			(0.090)	(0.093)
Nudge Effectiveness			0.049	0.009
			(0.101)	(0.104)
Nudge Painfulness			-0.110	0.005
			(0.099)	(0.104)
Conservative				-1.003^{***}
				(0.185)
(Intercept)	0.919^{***}	0.937^{***}	0.246	0.833
	(0.133)	(0.158)	(0.426)	(0.452)
Log Likelihood	-520.656	-520.630	-400.485	-385.410
Num. obs.	800	800	800	800

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Coefficients are expected log-odds.

Table A.3: Logistic regression for Study 2. Introducing a nudge aimed at increasing retirement savings into the choice set does not crowd-out support for the carbon tax. But, as before, introducing a green energy nudge leads to crowding-out. This finding holds when controlling for covariates.

	Model 1	Model 2	Model 3	Model 4
Tax + Nudge	-0.391^{**}	0.023	0.006	-0.025
	(0.146)	(0.201)	(0.249)	(0.253)
Related Nudge	-0.358^{*}	0.054	0.173	0.209
	(0.146)	(0.201)	(0.246)	(0.250)
Tax + Nudge x Related Nudge		-0.870^{**}	-1.275^{***}	-1.316^{***}
		(0.294)	(0.358)	(0.363)
Tax Effectiveness			0.913^{***}	0.870^{***}
			(0.095)	(0.096)
Tax Painfulness			-0.845^{***}	-0.819^{***}
			(0.083)	(0.084)
Conservative				-0.711^{***}
				(0.180)
(Intercept)	-0.047	-0.253	-0.360	0.042
	(0.124)	(0.143)	(0.385)	(0.404)
Log Likelihood	-532.411	-528.001	-389.545	-381.715
Num. obs.	802	802	802	802

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Coefficients are expected log-odds.

A.3 Study 2

We begin by looking at the effectiveness and painfulness of the three policies. On a five-point Likert scale, participants believed the green energy nudge to be more effective than the carbon tax (3.25 vs. 2.91, $t(401) = 6.36$, $p < .001$). Moreover, participants thought the retirement nudge was more effective at increasing retirement savings (3.65) than the green energy nudge was at reducing pollution ($t(798.16) = -5.92$, $p < .001$). Correspondingly, the carbon tax was viewed as more painful (3.22) than both the related nudge (1.71, $t(993.42) = -23.51$, $p < .001$) as well as the unrelated nudge (1.71, $t(1,001.01) = -23.76$, $p < .001$). There was no significant difference in how painfully the two nudges were rated ($t(799.89) = 0.06$, $p = .953$).

In Table A.3, we conduct the analysis in a regression framework. Model 1 shows the main effects of our manipulations. Support for implementing a carbon tax decreased both when we introduced a nudge into the choice set ($p < 0.01$) as well as when participants were presented with a related nudge ($p < 0.05$). We test for the predicted interaction in Model 2 and confirm that the crowding-out effect is entirely driven by the conditition in which the green energy nudge is introduced into the choice set ($p < 0.01$). Model 3 shows that this effect is robust to the controls we have used in previous studies. That is, participants who rated the carbon tax as more effective and who identify as liberal are both more likely to support its implementation (both $p < 0.001$). The interaction between the choice set and the related nudge remains significant at the same level ($p < 0.01$). Model 3 estimates that introducing a related nudge decreases the log-odds of supporting the tax by 60.5%, which is in line with the estimate from our previous studies.

A.4 Study 3A

Across our analyses, we follow our pre-registered plan to perform no comparisons across the two domains. Instead, we treat the environmental and retirement domains separately.

We begin by looking at the effectiveness and painfulness of the four policies. On a five-point Likert scale, participants believed the green energy nudge to be more effective at reducing carbon emissions and pollution (3.28) than the carbon tax (2.91, $t(609) = 7.29$, $p < .001$). At the same time, they thought the green energy nudge was less painful (1.67) than the carbon tax (3.01, $t(609) = -26.76$, $p < .001$).

In the retirement domain, participants believed defaulting employees into 401(k) plans to be more effective at promoting retirement savings (3.17) than increasing contributions to and benefits from social security (2.96, $t(597) = 3.66$, $p < .001$). At the same time, they thought the default was less painful (1.98) than expanding social security (2.86, $t(597) = -16.40$, $p < .001$).

It might be that merely having passed one policy diminishes support for a second policy. Just as reading about a nudge first reduced support for taxes, reading about taxes may have a similar effect on support for the nudge. Because all participants also made a decision about whether or not to implement the nudge, we can look at the effect of our conditions on that decision. We show

Figure A.1: Study 3A. The order in which the nudge is evaluated does not affect support for its implementation. That is, we observe no crowding-out for the nudge as we do for the tax. Error bars show $\pm$ one standard error.

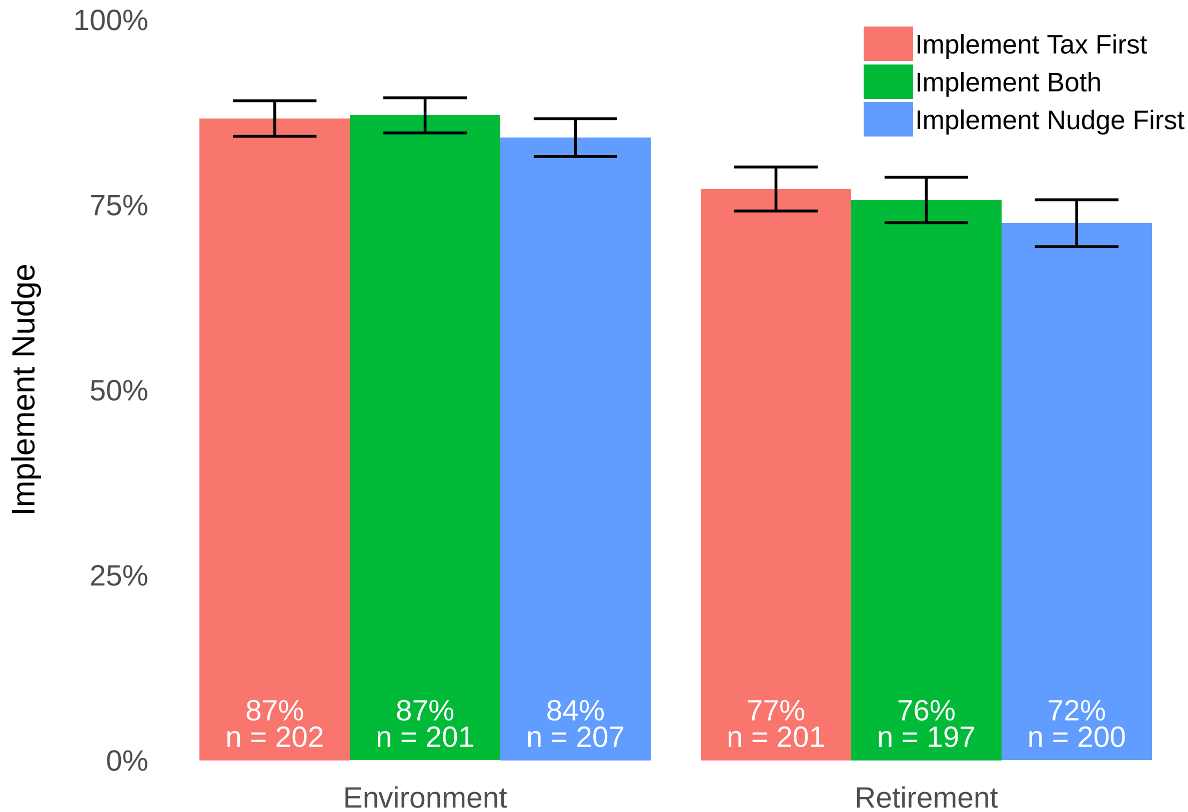


Table A.4: Logistic regression for the decision to implement the carbon (Models 1-3) and social security taxes (Models 4-6) in Study 3A. Support for implementing both a carbon tax and an increase in social security taxes decreases when participants make the decision either jointly with or following the decision to implement a nudge.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Implement Both	-0.632^{**}	-0.965^{***}	-0.975^{***}	-0.669^{**}	-1.139^{***}	-1.348^{***}
	(0.202)	(0.246)	(0.250)	(0.203)	(0.253)	(0.268)
Implement Nudge First	-0.745^{***}	-1.060^{***}	-1.066^{***}	-0.736^{***}	-1.116^{***}	-1.355^{***}
	(0.201)	(0.246)	(0.249)	(0.203)	(0.251)	(0.267)
Tax Effectiveness		0.852^{***}	0.823^{***}		1.176^{***}	1.209^{***}
		(0.105)	(0.109)		(0.119)	(0.129)
Tax Painfulness		-0.857^{***}	-0.805^{***}		-0.506^{***}	-0.658^{***}
		(0.102)	(0.105)		(0.104)	(0.115)
Nudge Effectiveness			-0.046			-0.284^{*}
			(0.108)			(0.112)
Nudge Painfulness			-0.037			0.397^{***}
			(0.120)			(0.108)
Conservative			-0.634^{**}			-0.611^{**}
			(0.202)			(0.212)
(Intercept)	0.443^{**}	0.724	1.191^{*}	0.372^{**}	-1.460^{**}	-0.576
	(0.144)	(0.444)	(0.573)	(0.144)	(0.499)	(0.625)
Log Likelihood	-414.760	-313.427	-308.172	-405.677	-300.229	-284.386
Num. obs.	610	610	610	598	598	598

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Coefficients are expected log-odds.

support for implementing the nudge (or both the tax and the nudge) in Figure A.1. F-tests confirm the graphic results in the figure: on average, 85.9% supported the green energy nudge and 85.9% supported the 401(k) contribution default, with no significant differences across the three conditions ($\chi^2(2, n = 610) = 0.89$, $p = .639$, in the environment domain and, $\chi^2(2, n = 598) = 1.19$, $p = .552$ in the retirement domain).

We again extend our analyses on the decision to implement the tax using logistic regression, shown in Table A.4. Our baseline Model 1 looks at the decision to implement the carbon tax. Confirming our previous analyses, we find that both the joint implementation decision and implementing the nudge first decrease the likelihood of supporting the tax ($p < 0.01$ and $p < 0.001$, respectively). Model 2 includes our previously used controls: the perceived effectiveness of the carbon tax at decreasing CO_2 emissions and pollution is associated with a higher likelihood of choosing to implement the tax, whereas a greater perceived painfulness is associated with a decrease (both $p < 0.001$). We find no effect on the implementation decision of how participants evaluated the nudge, while conservatives are again less likely to favor implementation of the carbon tax.

Models 3 and 4 perform the corresponding analyses in the retirement domain. We replicate all our findings in this domain, with a notable additional finding: when the decision involves increasing social security taxes, participants' perceived effectiveness and painfulness of the nudge does affect the decision to expand social security. In particular, those who thought a 401(k) default was more painful or less effective were also more likely to favor an expansion of social security.

We can further compare the coefficients for "Implement Both" and "Implement Nudge First" within each of the four models. Using a general linear hypothesis test, we cannot reject the null hypotheses that the coefficients are identical (all $p > 0.60$). Making the joint implementation decision therefore appears to induce the same amount of crowding-out as does the decision to implement the nudge first.

Finally, we perform the same regression analyses with the decision to implement the nudge (or both the tax and the nudge) in Table A.5. As suggested by our previous analyses and apparent in Figure A.1, our experimental conditions do not affect the decision to implement the nudge. Participants who believe the nudge to be more effective and less painful are more likely to favor its implementation. We observe that for the politicized environmental domain, conservatives were also less likely to support implementing a nudge. Conversely, politicial affiliation is not associated with support for the retirement savings nudge.

A.5 Study 3B

We relied on a Latin Squares design in the expectation that there would be no spillover across the two domains. That is, having seen a nudge in the domain of environment should not affect the decision to implement a tax in the retirement domain. We test for such a spillover using a

Table A.5: Logistic regression for the decision to implement the nudges in Study 3A (Models 1-3: carbon tax; Models 4-6: 401(k)). While perceived nudge effectiveness and painfulness affect the decision to implement the corresponding nudge, the ordering or joint implementation manipulations do not.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Implement Both	0.038	0.004	0.145	-0.082	-0.091	-0.036
	(0.295)	(0.300)	(0.349)	(0.236)	(0.237)	(0.270)
Implement Nudge First	-0.206	-0.265	-0.157	-0.245	-0.252	-0.243
	(0.281)	(0.288)	(0.327)	(0.231)	(0.232)	(0.266)
Tax Effectiveness		0.435^{***}	0.078		0.063	-0.079
		(0.119)	(0.136)		(0.093)	(0.113)
Tax Painfulness		-0.109	0.145		-0.072	0.091
		(0.110)	(0.136)		(0.093)	(0.114)
Nudge Effectiveness			1.027^{***}			1.078^{***}
			(0.157)			(0.125)
Nudge Painfulness			-0.746^{***}			-0.533^{***}
			(0.145)			(0.105)
Conservative			-1.184^{***}			-0.392
			(0.321)			(0.225)
(Intercept)	1.869^{***}	1.055	0.368	1.215^{***}	1.241^{*}	-0.645
	(0.207)	(0.554)	(0.743)	(0.168)	(0.485)	(0.646)
Log Likelihood	-247.674	-238.186	-176.858	-335.135	-334.273	-265.605
Num. obs.	610	610	610	598	598	598

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Coefficients are expected log-odds.

Table A.6: Mixed effects logistic regression for the decision to implement the tax (Model 1) and Nudge (Model 2) in Study 3B. We observe that the tax is less likely to be implemented when it is shown first in the second domain, consistent with a spillover effect in our Latin Squares design. We consequently limit our analysis to only the first domain participants had encountered.

	Model 1	Model 2	Model 3	Model 4
Retirement Domain	-0.455^{***}	-0.456^{***}	0.059	0.059
	(0.130)	(0.130)	(0.184)	(0.184)
Implement Tax First	0.358^{**}	0.724^{***}	-0.235	-0.117
	(0.130)	(0.202)	(0.185)	(0.311)
Second Domain	-0.072	0.281	0.045	0.168
	(0.128)	(0.194)	(0.184)	(0.319)
Implement Tax First x Second Domain		-0.726^{*}		-0.237
		(0.300)		(0.505)
(Intercept)	0.730^{***}	0.561^{***}	2.885^{***}	2.823^{***}
	(0.138)	(0.152)	(0.432)	(0.442)
Log Likelihood	-826.914	-823.953	-525.975	-525.864
Num. obs.	1282	1282	1282	1282
Num. groups: id	641	641	641	641
Var: id (Intercept)	0.913	0.895	3.916	3.899

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Coefficients are expected log-odds.

mixed-effects logistic regression with fixed effects for the domain (retirement or environment), a dummy variable that is equal to 1 if the tax was shown first, a dummy variable that is equal to 1 if the response is in the second domain, and the interaction between those two variables. We also include a random effect at the individual level, accounting for the fact that participants made two decisions and those may correlate. We find a significant interaction between the two dummies in the decision to implement the nudge (shown in Table A.6). That is, respondents are less likely to implement a tax when it is shown first in the second domain. Because this suggests spillover, and consistent with our pre-registered analysis plan, we limit our analysis to only the first domain participants had encountered.

We conduct a series of logistic regressions on the decision to implement the environment and retirement taxes, shown in Table A.7. The dependent variable for Models 1 and 2 is the decision to implement the carbon tax. We observe that first making the decision to implement the nudge makes it less likely for the tax to be implemented, both without controls in baseline Model 1 ($p < 0.01$)

Table A.7: Logistic regression for the decision to implement the carbon tax (Models 1 and 2) and the expanded social security tax (Models 3 and 4) in Study 3B. We observe that the carbon tax is less likely to be implemented when the decision to implement the nudge is made first, but do not observe significant crowding-out for the social security tax. Model 5 combines both domains.

	Model 1	Model 2	Model 3	Model 4	Model 5
Implement Nudge First	-0.802^{**}	-0.758^{*}	-0.418	-0.480	-0.569^{**}
	(0.245)	(0.307)	(0.230)	(0.267)	(0.198)
Tax Effectiveness		1.101^{***}		0.459^{***}	0.671^{***}
		(0.168)		(0.108)	(0.090)
Tax Painfulness		-0.821^{***}		-0.883^{***}	-0.848^{***}
		(0.166)		(0.148)	(0.108)
Conservative		-0.970^{**}		-0.676^{*}	-0.791^{***}
		(0.301)		(0.267)	(0.197)
Retirement Domain					-0.594^{**}
					(0.199)
(Intercept)	1.166^{***}	0.752	0.594^{***}	1.805^{***}	1.825^{***}
	(0.186)	(0.621)	(0.171)	(0.520)	(0.418)
Log Likelihood	-196.680	-141.233	-214.855	-171.530	-318.908
Num. obs.	321	321	320	320	641

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Coefficients are expected log-odds.

and with controls in Model 2 ($p < 0.05$). In Models 3 and 4, we perform the corresponding analysis for the decision to implement the expanded social security tax. Contrary to Study 3A, we do not observe significant crowding-out here. Although our predictions hold directionally, they do not reach conventional levels of significance (p = 0.07 with and without controls). In Model 5, we combine both of our domains and include a domain control variable. In the combined data, we again observe the hypothesized crowding-out effect ($p < 0.01$).

We conclude the analysis of our full data with a look at whether the tax may conversely have crowded out suport for the nudge. We show the regressions in Table A.8, with the format following that of the previous table for the tax. That is, Models 1 and 2 look at the decision to implement the green energy nudge, Models 3 and 4 look at the decision to implement the 401(k) contribution nudge, and Model 5 combining both domains. As predicted, we observe no crowding-out when it comes to the decision to implement the nudge in either domain, with and without controls.

We find support for the nudge in the environment domain ranging from 84.06% to 87.06%

Table A.8: Logistic regression for the decision to implement the carbon nudge (Models 1 and 2) and the retirement savings nudge (Models 3 and 4) in Study 3B. Model 5 combines both domains. Across all models, we observe no crowding-out of the nudge.

	Model 1	Model 2	Model 3	Model 4	Model 5
Implement Nudge First	-0.098	-0.256	0.255	0.148	-0.014
	(0.325)	(0.435)	(0.299)	(0.333)	(0.261)
Tax Effectiveness		1.539^{***}		1.024^{***}	1.211^{***}
		(0.290)		(0.179)	(0.150)
Tax Painfulness		-1.066^{***}		-0.428^{*}	-0.673^{***}
		(0.261)		(0.183)	(0.147)
Conservative		-2.023^{***}		-0.708^{*}	-1.250^{***}
		(0.489)		(0.335)	(0.268)
Retirement Domain					-0.579^{*}
					(0.265)
(Intercept)	1.890^{***}	0.959	1.464^{***}	-0.491	0.380
	(0.234)	(0.869)	(0.210)	(0.640)	(0.518)
Log Likelihood	-128.230	-72.352	-144.883	-118.015	-196.250
Num. obs.	321	321	320	320	641

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Coefficients are expected log-odds.

($F(2, 607) = 0.45$, $MSE = 0.12$, $p = .640$) and in the retirement domain ranging from 72.5% to 77.11% ($F(2, 595) = 0.59$, $MSE = 0.19$, $p = .553$).

A.5.0.1 Policymakers

As pre-registered, we repeat the analysis for the subset of participants (n = 347) who report active involvement in either impacting or informing public policy. We parallel the analysis for the full set of respondents.

Policymakers evaluated the nudge as no more (or less) effective than a carbon tax at reducing CO_2 emissions and pollution (3.18 vs 3.06, $t(347.23) = 1.12$, $p = .265$). In the retirement domain, policymakers, much like the full sample, believed a 401(k) nudge to be more effective than an expansion of social security (3.33 vs 2.99, $t(302.87) = 2.73$, $p = .007$). Similar to our previous respondents, they also view the nudges as less painful than the taxes in both domains (1.66 vs 2.75 in environment, $t(338.43) = -11.26$, $p < .001$; 1.54 vs 2.44 in retirement $t(322.18) = -8.55$, $p < .001$).

Figure A.2: Study 3B subset of policymakers. Even among this more expert sample, we observe crowding-out in support for the tax (left panel), but not for the nudge (right panel). Error bars show ± one standard error.

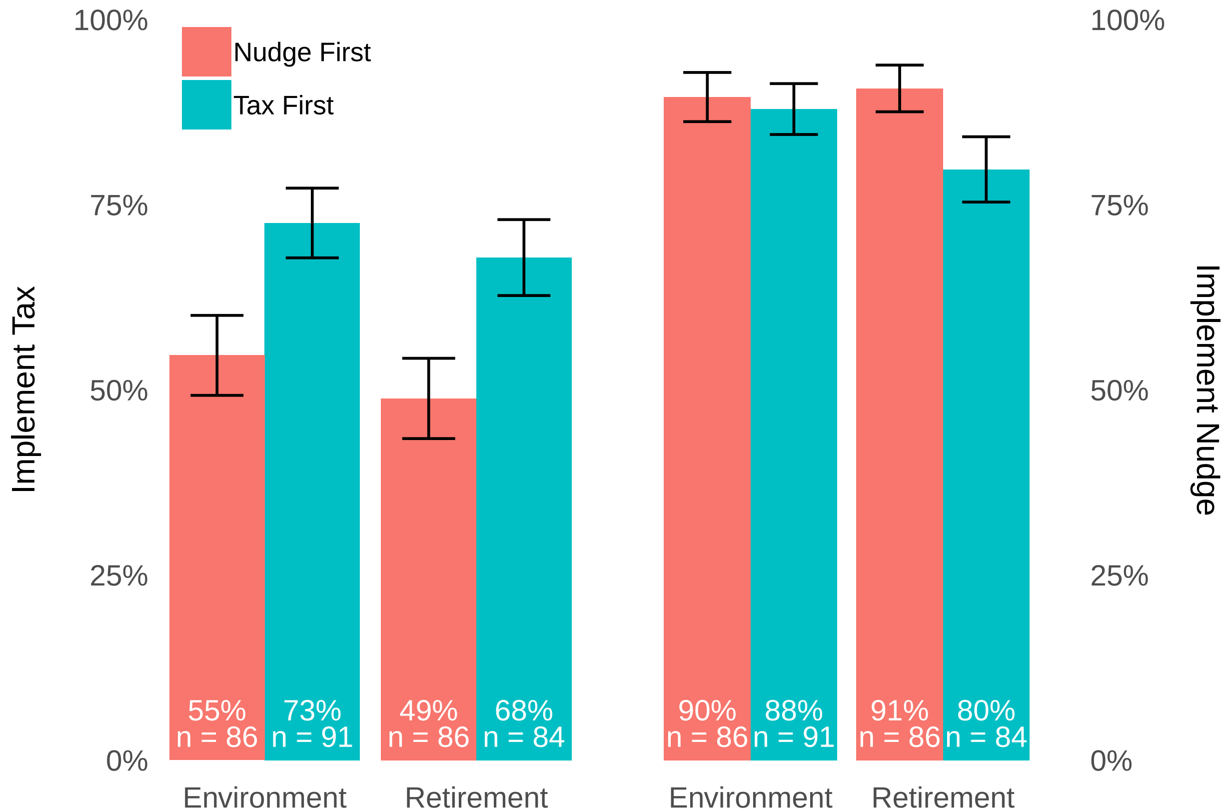


Table A.9: Logistic regression for policymakers' decisions to implement the carbon tax (Models 1 and 2) and the expanded social security tax (Models 3 and 4) in Study 3B. We observe that both taxes are less likely to be implemented without controls (Models 1 and 3), but the effect is not significant with controls (Models 2 and 4). When we combine both domains (Model 5), crowding-out again reaches the conventional level of significance.

	Model 1	Model 2	Model 3	Model 4	Model 5
Implement Nudge First	-0.784^{*}	-0.777	-0.794^{*}	-0.592	-0.590^{*}
	(0.319)	(0.429)	(0.318)	(0.367)	(0.272)
Tax Effectiveness		1.362^{***}		0.509^{***}	0.790^{***}
		(0.230)		(0.144)	(0.120)
Tax Painfulness		-0.636^{**}		-0.901^{***}	-0.796^{***}
		(0.221)		(0.203)	(0.145)
Conservative		-0.842^{*}		-0.465	-0.659^{*}
		(0.422)		(0.368)	(0.271)
Retirement Domain					-0.409
					(0.274)
(Intercept)	0.971^{***}	-0.879	0.747^{**}	1.573^{*}	1.026
	(0.235)	(0.861)	(0.234)	(0.645)	(0.533)
Log Likelihood	-112.737	-74.909	-112.334	-90.379	-171.990
Num. obs.	177	177	170	170	347

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Coefficients are expected log-odds.

Figure A.2 shows support for the tax in the domains of environment and retirement (left panel), and the support for the nudge (right panel), conditional on making the decision to implement a nudge first (red bar) or the tax first (blue bar). In both domains, we see that presenting the nudge first reduces support for the tax. In the environment domain, support declines from 72.53% to 54.65% ($\chi^2(1, n = 177) = 5.37$, $p = .020$); in the retirement domain, we see (in contrast to the full sample) also a significant decline from 67.86% to 48.84% ($\chi^2(1, n = 170) = 5.56$, $p = .018$).

Consistent with the results from the previous study, we find no crowding-out in support for the nudge. In the environment domain, 87.91% support implementing the nudge when it is made prior to the decision to implement the tax and 89.53% do so when it follows that decision ($\chi^2(1, n = 177) = 0.01$, $p = .918$). In the retirement domain, we similarly see no decrease with 79.76% supporting its implementation when asked first and 90.7% do so when asked second ($\chi^2(1, n = 170) = 3.23$, $p = .072$).

Table A.10: Logistic regression for policymakers' decision to implement the carbon nudge (Models 1 and 2) and the retirement savings nudge (Models 3 and 4) in Study 3B. Model 5 combines both domains. We find some crowding out of the retirement savings nudge in Model 3 that disappears with controls and in the combined model.

	Model 1	Model 2	Model 3	Model 4	Model 5
Implement Nudge First	0.162	-0.470	0.906^{*}	0.813	0.369
	(0.477)	(0.696)	(0.460)	(0.520)	(0.397)
Tax Effectiveness		2.268^{***}		1.297^{***}	1.553^{***}
		(0.553)		(0.303)	(0.250)
Tax Painfulness		-1.317^{**}		-0.205	-0.534^{*}
		(0.489)		(0.264)	(0.215)
Conservative		-1.787^{*}		-0.968	-1.306^{**}
		(0.729)		(0.513)	(0.405)
Retirement Domain					-0.722
					(0.403)
(Intercept)	1.984^{***}	-0.035	1.371^{***}	-1.598	-0.597
	(0.322)	(1.290)	(0.272)	(1.051)	(0.790)
Log Likelihood	-62.375	-30.487	-68.925	-52.391	-87.462
Num. obs.	177	177	170	170	347

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Coefficients are expected log-odds.

We next conduct a serious of logistic regressions on the decision to implement the environment and retirement taxes, shown in Table A.9. The dependent variable for Models 1 and 2 is the decision to implement the carbon tax. We observe in Model 1 that, without additional controls, first making the decision to implement the nudge makes it less likely for the tax to be implemented ($p < 0.05$). When we add controls in Model 2, however, this effect is no longer significant (p = 0.07). We similarly observe crowding-out for the social security tax without further controls (Model 3, $p < 0.05$), but not with controls (Model 4; p = 0.11). In Model 5, we combine both of our domains and include a domain control variable. In the combined data, we again observe the hypothesized crowding-out effect ($p < 0.05$).

We conclude the analysis of the policymaker respondents with a look at whether the tax may have crowded out support for the nudge. We show the regressions in Table A.10, with the format following that of the previous table for the tax. That is, Models 1 and 2 look at the decision to implement the green energy nudge, Models 3 and 4 look at the decision to implement the 401(k) contribution nudge, and Model 5 combining both domains. Against our predictions and previous results, we do observe some crowding out of the retirement savings nudge without controls (Model 3). This effect again disappears with controls (Model 4, $p = 0.12$) and when we combine both domains (Model 5, $p = 0.35$).

As before, we observe no comparable crowding-out for the nudge in either the environmental domain ($F(1, 319) = 0.09$, $MSE = 0.12$, $p = .763$) or the retirement domain ($F(1, 318) = 0.73$, $MSE = 0.14$, $p = .394$). That is, respondents appear to be willing to implement two policies, just as long as the second option offered to them is a (painless) nudge rather than a (painful) tax.

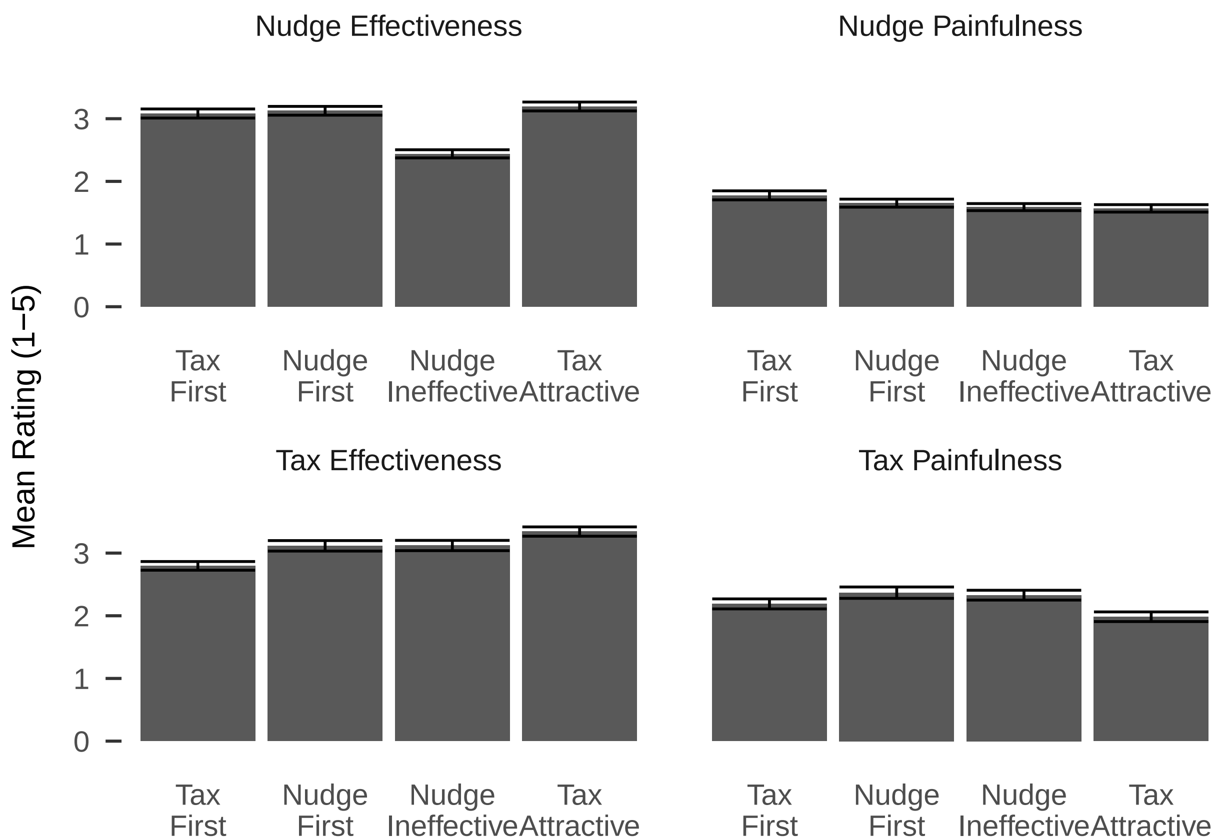
A.6 Study 4

We begin by testing the effect of our two experimental manipulations. The mean ratings for effectiveness and painfulness of both policies and for each condition is shown in Figure A.3. First, we look at whether providing additional information in the "Nudge Ineffective" condition lowers participants' rating of the nudge. Indeed, we find that with information about the nudge's effect size, perceived effectiveness declines from 3.13 ("Nudge First") to 2.44 ($t(389.61) = 7.17$, $p < .001$). Notably, there is no change in the perceived painfulness of the nudge (1.65 and 1.59, respectively; $t(383.32) = 0.74$, $p = .458$)

When we make the tax more attractive, by highlighting that the funds could be used to offset other taxes and promote investment, participants rated it as less painful (1.98) than in the "Nudge First" condition (2.37), which featured the identical ordering of questions but with this information omitted ($t(371.59) = 3.22$, $p = .001$). Participants also rated such a tax as more effective at reducing emissions (3.34) than when the tax was introduced without this additional information (3.34, $t(376.30) = -2.04$, $p = .042$).

We next return to our logistic regression analysis, shown in Table A.11. We use as baseline the "Nudge First" condition, in which we observed the least support for the carbon tax. In Model 1,

Figure A.3: Study 4: Ratings for painfulness and effectiveness of a green energy nudge and a carbon tax across conditions. Error bars show ± one standard error.



we regress dummy variables for the conditions on the decision to implement the tax. Consistent with the chi-square regression reported in the manuscript, ordering on its own did not significantly change support for the tax. However, framing either the nudge as ineffective or the carbon tax as being less painful increases support for the tax. Model 2 includes controls for how effective participants rated the carbon tax and the green energy nudge. Consistent with our previous findings, rating the tax as more attractive and less painful correspond with increased support for implementing it (both $p < 0.001$). With these controls included, we now see an effect of ordering ("Tax First") as well as framing the nudge as ineffective (both $p < 0.05$), with the effect of the tax attractive manipulation absorbed in the effect of painfulness. Finally, Model 3 includes a control for political affiliation. Conservatives are less likely to support the tax ($p < 0.001$), but the coefficients on the other variables remain nearly unchanged.

We also asked participants how they thought their own carbon emissions compared to the average household, whether they thought human activity or nature was primarily driving global climate change, and whether they thought the government should intervene in reducing carbon emissions. Although we preregistered no hypotheses, we report descriptively how these groups differed in their support for the tax and the nudge.

53.88% of respondants thought their emissions were about average and 43.86% thought they were less than average. Surprisingly, only 18 participants thought they polluted more than average. Support for implementing the tax did not differ by one's own emissions $F(2, 795) = 0.49$, $MSE = 0.22$, $p = .612$. We have no reason to believe our sample to be more environmentally conscious than the population on average. This suggests that per-capita refunds of carbon taxes might be perceived by most people as either neutral on their household budget or even generate income for them. Future work may want to examine whether highlighting the redistributive impact of a carbon tax might enhance its acceptance.

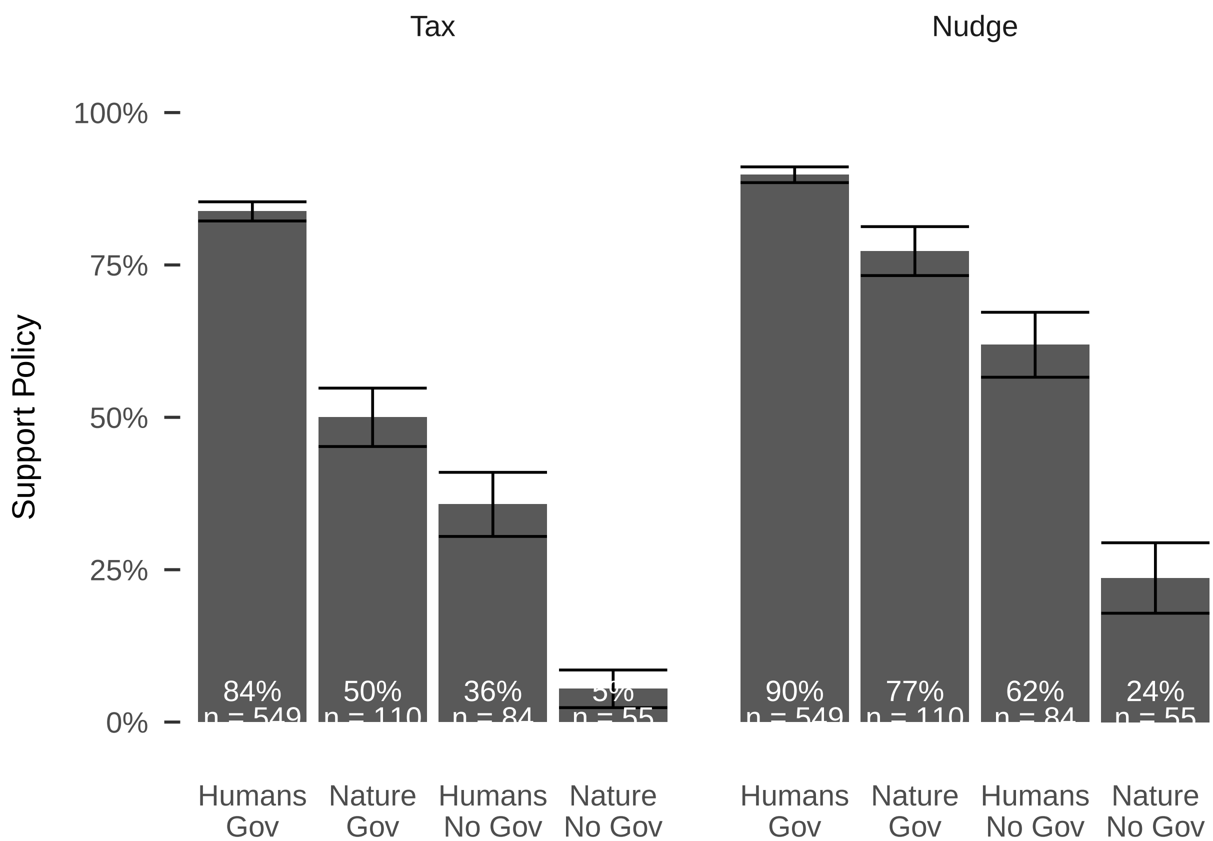
Participants who favored government intervention to reduce carbon emissions were, not surprisingly, more supportive of the carbon tax (see left panel of Figure A.4). Support was greatest among the

Table A.11: Logistic regression for the decision to implement the carbon tax in Study 4. Highlighting the small effect size of the nudge increases support for the tax, as does making the tax appear more attractive and less painful (Model 1). Including controls for effectiveness and painfulness of the policies also replicates the finding that merely asking first about the nudge can reduce support for the tax.

	Model 1	Model 2	Model 3
Tax First	0.325	0.712^{*}	0.611^{*}
	(0.212)	(0.285)	(0.295)
Nudge Ineffective	0.424^{*}	0.709^{*}	0.685^{*}
	(0.214)	(0.299)	(0.311)
Tax Attractive	0.514^{*}	0.056	0.011
	(0.219)	(0.298)	(0.306)
Tax Effectiveness		1.275^{***}	1.224^{***}
		(0.121)	(0.124)
Tax Painfulness		-1.173^{***}	-1.167^{***}
		(0.112)	(0.116)
Nudge Effectiveness		-0.053	-0.113
		(0.112)	(0.116)
Nudge Painfulness		-0.129	-0.069
		(0.115)	(0.121)
Conservative			-1.217^{***}
			(0.217)
(Intercept)	0.472^{**}	-0.137	0.781
	(0.149)	(0.500)	(0.545)
Log Likelihood	-492.904	-303.141	-286.727
Num. obs.	798	798	798

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Coefficients are expected log-odds.

Figure A.4: Support for the carbon tax and green energy nudge in Study 4 for participants who thought human activity or natural forces were primarily responsible for climate change and who thought the government should or should not intervene to reduce emissions.



68.8% who also believed human activity to be the primary driver of global climate change: of those, 83.79% supported the carbon tax. Of the 10.53% of participants who thought the government should not take any action even as humans were primarily responsible, only 35.71% supported the carbon tax. Fewer participants believed nature to be the primary driver of warming. Notably, even among this group, more participants favored the government taking action to limit human contribution (13.78%) than did not want government to be involved (6.89% of the total sample). All pairwise Chi-square tests were significant ($p < 0.01$) except for the comparison between Humans + No Gov and Nature + Gov ($p = 0.07$). The pattern holds similarly for the nudge (see right panel of Figure A.4). Chi-squared tests of all pair-wise comparisons are significant (for Humans + No Gov vs. Nature + Gov, $p = 0.03$, all other $p < 0.01$).

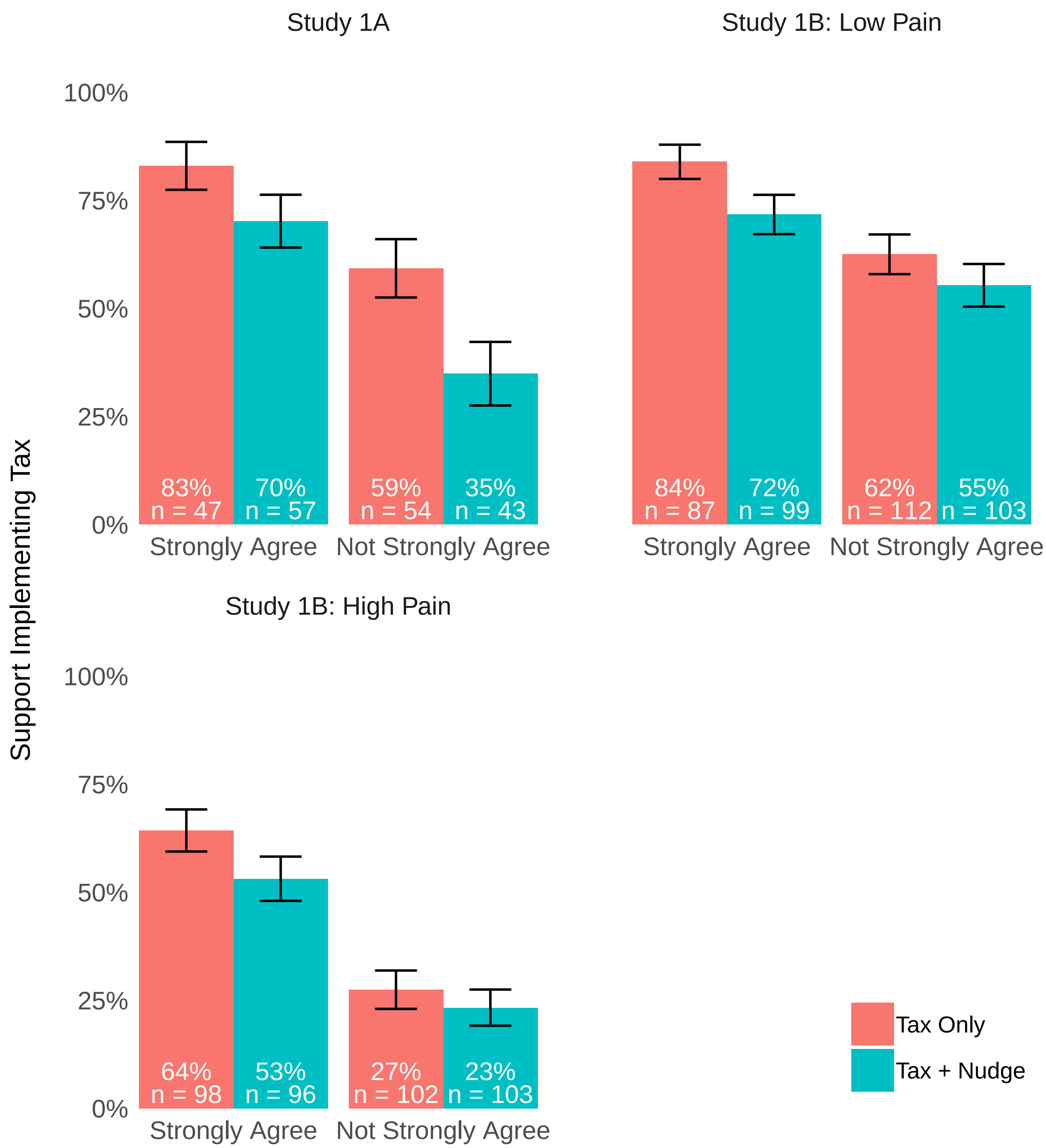
A.7 Heterogeneous Treatment Effects

Across our studies, we collected information about participants' political orientation, as well as how effective they believed the green energy nudge and the carbon tax to be. Moreover, in Studies 1A and 1B, we also asked participants about their belief in climate change. Although these analyses were not planned, we can pool responses from all our Amazon Mechanical Turk participants and explore whether effects differ consistently for some group of participants.

In Studies 1A and 1B, we elicited belief in the existence of climate change on a scale from 1 (strongly disagree) to 5 (strongly agree). Combining across these studies, we had 1001 participants. Of those, the majority (n = 484) strongly agreed that global average temperatures had been increasing over the past 50 years. Nearly as many participants somewhat agreed with that statement (n = 386), while the remaining 131 participants were either uncertain or somewhat or strongly disagreed.

In Figure A.5, we show support for implementing the carbon tax in our two experimental conditions for those who strongly agreed with the statement that temperatures have been increasing and the remaining participants. Note that the experimental design of "Study 1A" and "Study 1B: Low Pain" were identical and "Study 1B: High Pain" explicitly highlighted some of the costs of a carbon tax

Figure A.5: Support for implementing the carbon tax for those who strongly agree that climate change is occuring vs. remaining participants (Studies 1A and 1B). Crowding-out occurs even for those who expressed strong agreement. Error bars show ± one standard error.



for consumers. We observe a clear main effect, with those who are in strong agreement that climate change is occuring also more supportive of implementing a carbon tax.

To test for an interaction, we rely on a logistic regression and a new method for assessing nonlinear interaction effects (Hainmueller, Mummolo, and Xu, n.d.). If crowding-out differs by agreement with the existence of climate change, then we should observe a significant interaction effect between our experimental assignment and the belief in climate change. The first column of Table A.12 shows a logistic regression on the decision to support implementing the carbon tax. As predictors, we include dummy variables controlling for each of the studies as well as the different framings in Study 1B. Notably, the interaction between our experimental condition and belief in climate change is not significant, suggesting that crowding-out is not less pronounced among those who strongly agree that climate change is occuring. Moreover, the statistical result is robust to using a cutoff of "somewhat agree" instead and for treating the scale response as a continuous variable.

We also test for nonlinear interaction effects using a new approach, which bins the moderator variable (in this specification, belief in climate change occurring) into more granular cutoffs to assess for nonlinear marginal effects.(Hainmueller, Mummolo, and Xu, n.d.) We choose a four-fold partition (of those who (1) either disagreed or strongly disagreed, (2) neither agreed nor disagreed and (3) agreed, and (4) strongly agreed), respectively), which yielded a non-significant result for the test of nonlinear marginal effects, $p = 0.21$, using a Wald test (see Figure A.6).

In all our studies, we asked participants (n = 1806) about their ideological orientation, ranging from extremely liberal to extremely conservative. We perform a median split on political orientation (extremely conservative to extremely liberal) and show support for the carbon tax across all studies for the two groups in Figure A.7. The corresponding regression analysis is in column 2 of Table A.12.

Although we observe a main effect of political affiliation, with conservatives less supportive of a carbon tax, we do not observe a difference in our experimental treatment. That is, both liberals and conservatives appear equally discouraged from implementing a carbon tax by the presence of a

Table A.12: Analysis of heterogeneous treatment effects. The first column looks at belief in climate change, combining data from Studies 1A and 1B. Columns 2-4 combine data from Studies 1A, 1B, 2, and 3A (all experiments performed on Amazon Mechanical Turk) and look at differential effects by political orientation (column 3), perceived effectiveness of the carbon tax (column 4), and effectiveness of the carbon tax relative to the green energy nudge (column 5). We observe that there is less crowding-out among those who perceive the tax to be ineffective, where there is little support for implementing the tax in the baseline condition.

	CC Belief	Politics	Tax Effectiveness	Relative Effectiveness
Study 1B: Low Pain	0.30	0.39^{*}	0.23	0.05
	(0.19)	(0.19)	(0.19)	(0.04)
Study 1B: High Pain	-0.91^{***}	-0.82^{***}	-0.93^{***}	-0.20^{***}
	(0.19)	(0.19)	(0.19)	(0.04)
Study 2: Related Nudge		-1.06^{***}	-1.12^{***}	-0.24^{***}
		(0.19)	(0.19)	(0.04)
Study 3A: Environment		-0.31	-0.33	-0.07
		(0.19)	(0.19)	(0.04)
Tax + Nudge	-0.41^{*}	-0.76^{***}	-1.06^{***}	-0.15^{***}
	(0.19)	(0.15)	(0.20)	(0.04)
Strongly Agree with CC	1.32^{***}			
	(0.20)			
Conservative		-1.36^{***}		
		(0.15)		
Tax Ineffective			-1.74^{***}	
			(0.18)	
Nudge More Effective				-0.12^{**}
				(0.04)
Tax More Effective				0.15^{***}
				(0.04)
Tax + Nudge x Strongly Agree with CC	-0.21			
	(0.28)			
Tax + Nudge x Conservative		0.24		
		(0.21)		
Tax + Nudge x Tax Ineffective			0.57^{*}	
			(0.23)	
Tax + Nudge x Nudge More Effective				0.02
				(0.05)
Tax + Nudge x Tax More Effective				0.03
				(0.06)
(Intercept)	0.15	1.46^{***}	2.05^{***}	0.69^{***}
	(0.18)	(0.18)	(0.22)	(0.04)
Log Likelihood	-611.75	-1104.64	-1092.63	-1192.46
Num. obs.	1001	1806	1806	1806

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Coefficients are expected log-odds.

Figure A.6: Support for carbon tax by condition across four interaction variables and binning sizes.

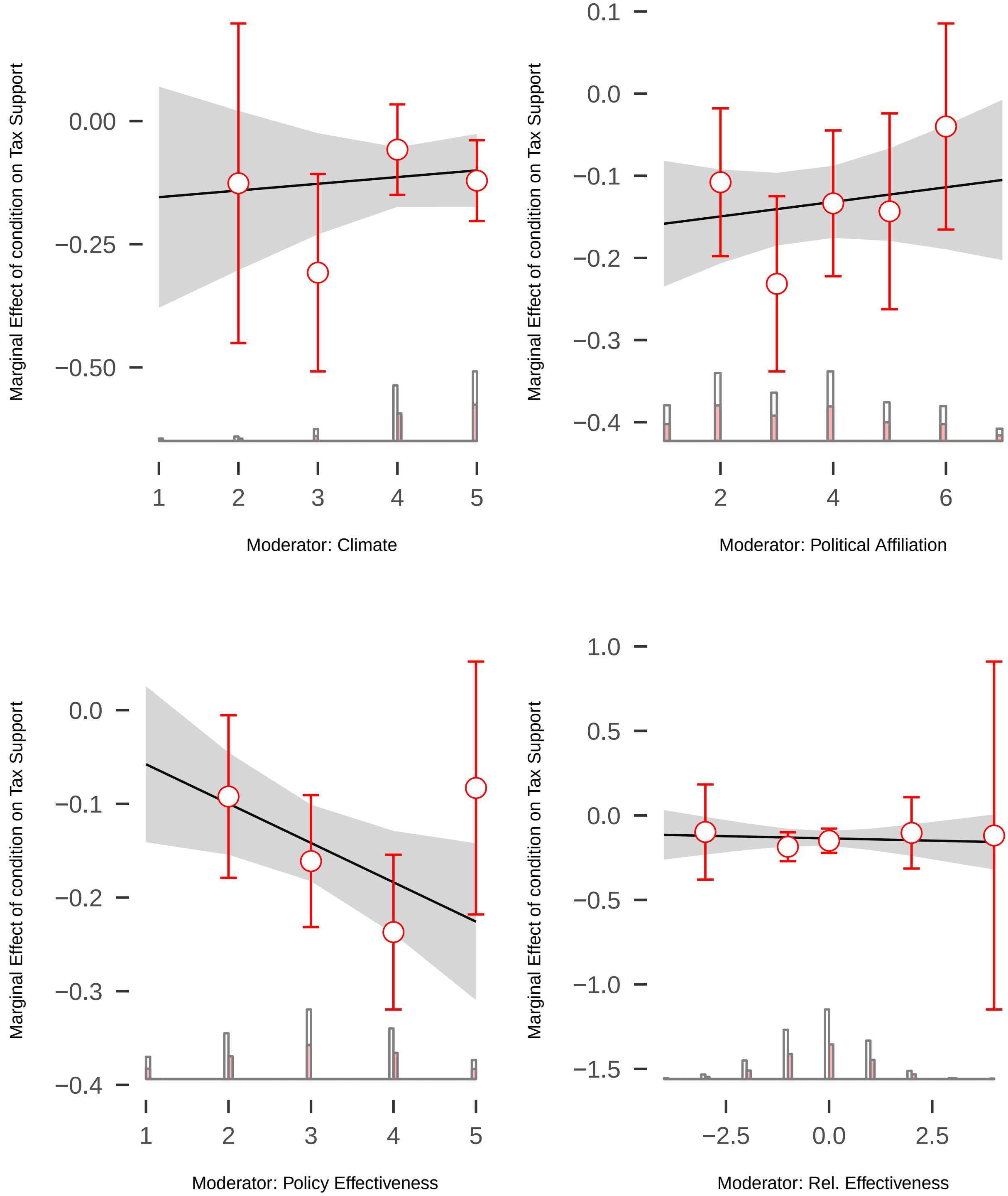
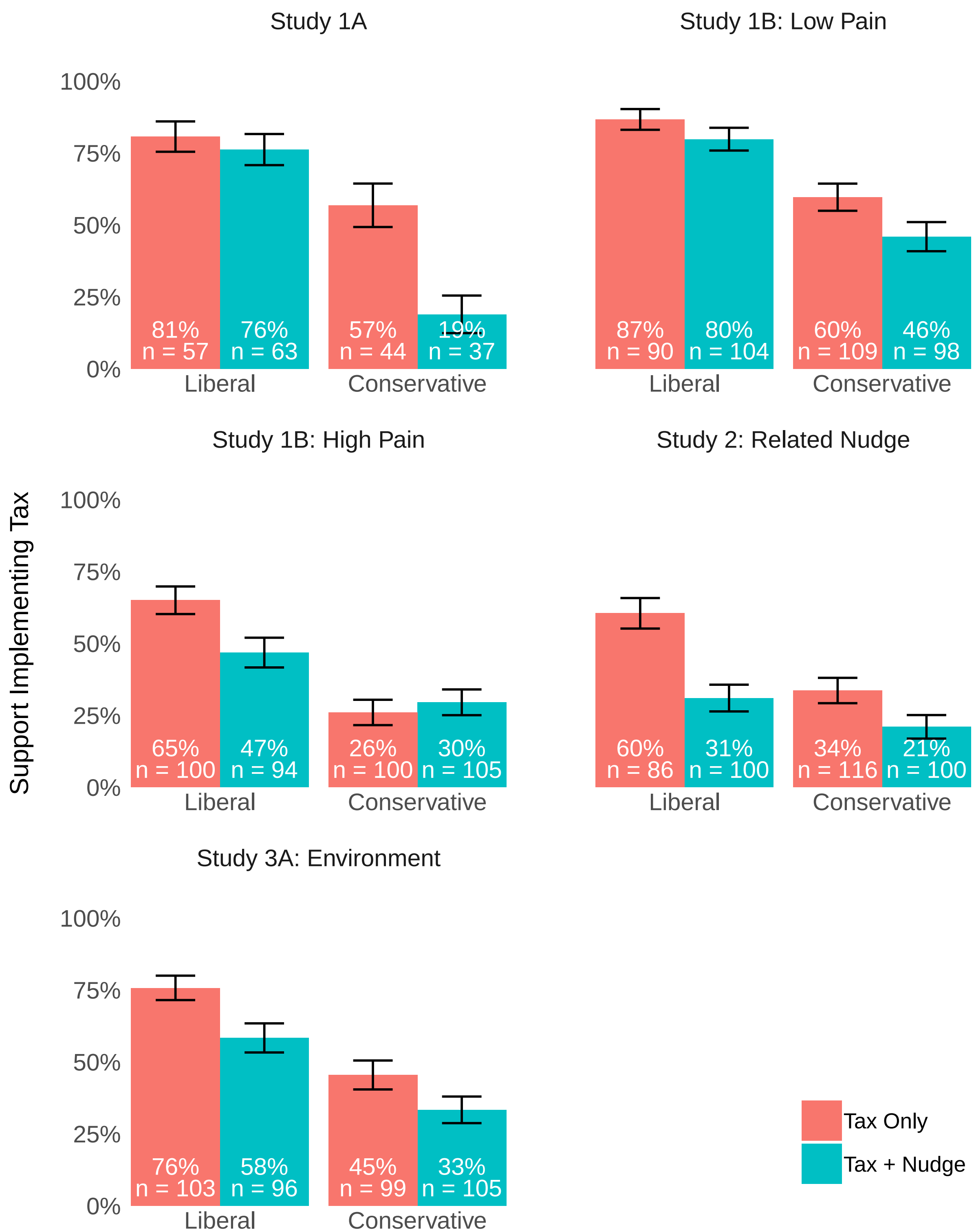


Figure A.7: Support for carbon tax by condition and by political affiliation for all studies with a joint implementation condition. Although conservatives are overall less likely to support implementing the tax, the extent of crowding-out by introducing the nudge does not differ by political affiliation. Error bars show $\pm$ one standard error.



green energy nudge. This result is robust to using different cutoff points to split conservatives and liberals and to treating politicial orientation as a continuous variable, and is confirmed by a Wald test, $p = 0.8$, indicating the four-bin model and the linear interaction model are not statistically different from each other.

It might also be that participants who believe the tax to be highly effective would be less affected by the introduction of a green energy nudge. Participants rated the perceived effectiveness on a five-point scale from "not effective at all" to "extremely effective." The most common response was the midpoint of the scale, "moderately effective" (n = 606), with 593 participants believing the tax to be less effective than that and 607 believing it to be more effective. We split participants according to whether they thought the tax was very or extremely effective ("Tax Effective") or whether they gave a lower response on the scale ("Tax Ineffective").

In Figure A.8 and in column 3 of Table A.12, we show support for implementing the carbon tax based on the perceived effectiveness of the tax. Not surprisingly, those who believe the tax to be ineffective are less supportive of implementing it. However, we also observe an interaction with our treatment – but in the opposite direction that might have been expected. Participants who believe the tax to be ineffective are affected less by the presence of a nudge. The coefficient on the interaction remains significant if we treat perceived effectiveness of the carbon tax as a continuous variable or if we choose a lower cutoff point (below "moderately effective") instead. A possible explanation for this finding is that support for the carbon tax is already low in the absence of a nudge and consequently we may observe a floor effect. A more nuanced interaction model that separates the tax's perceived effectiveness into four bins shows a non-significant effect, $p = 0.38$.

Finally, we can look at support for implementing the carbon tax as a function of whether participants thought the tax to be less effective than the nudge, equally effective, or more effective. We show this in Figure A.9 and column 4 of Table A.12. We again observe a main effect across studies, with those who believe the tax to be relatively less effective also less supportive of implementing it. However, we again observe no interaction with our experimental treatment. Testing a four-bin

Figure A.8: Support for carbon tax by condition and by perceived effectiveness of the tax for all studies with a joint implementation condition.

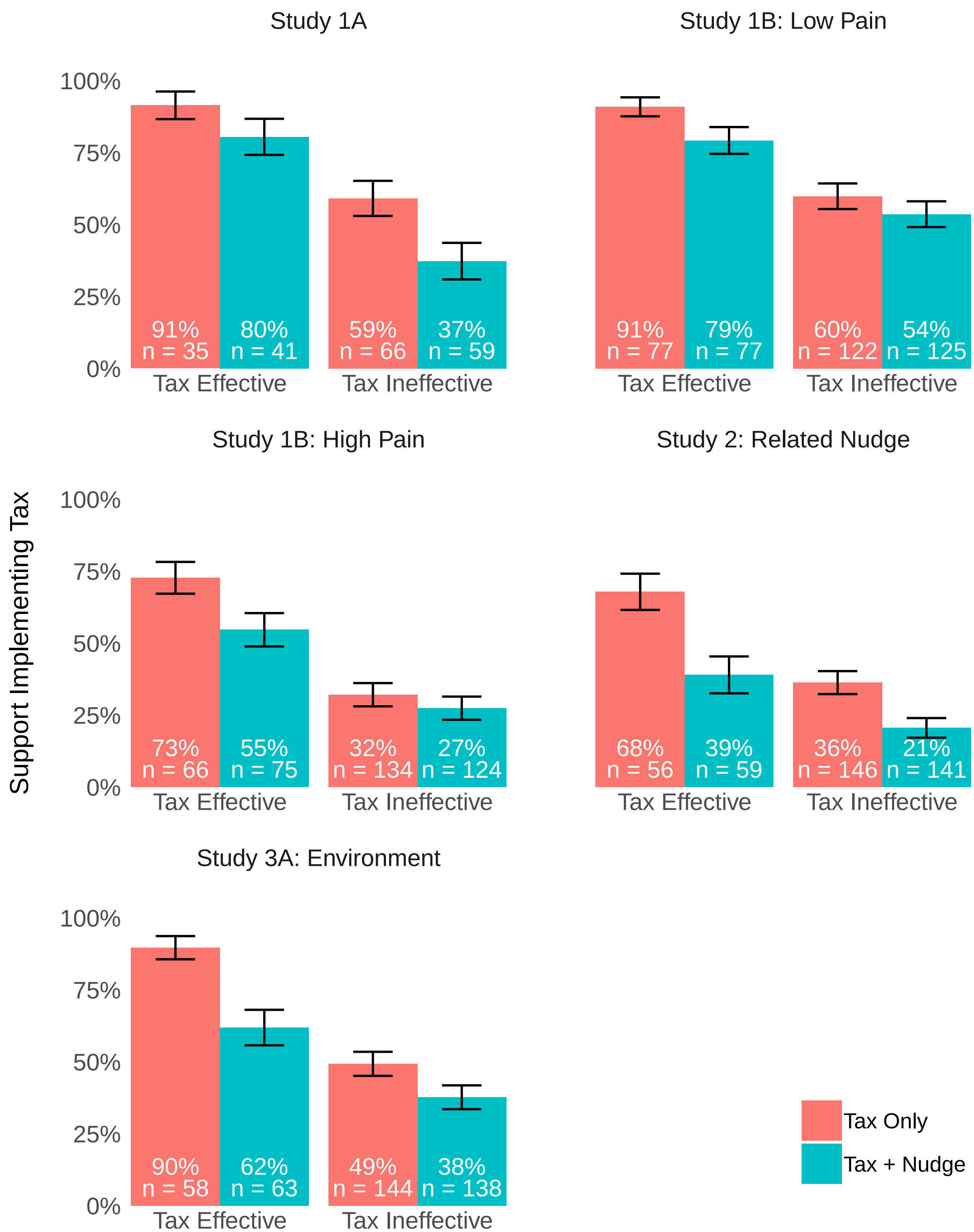
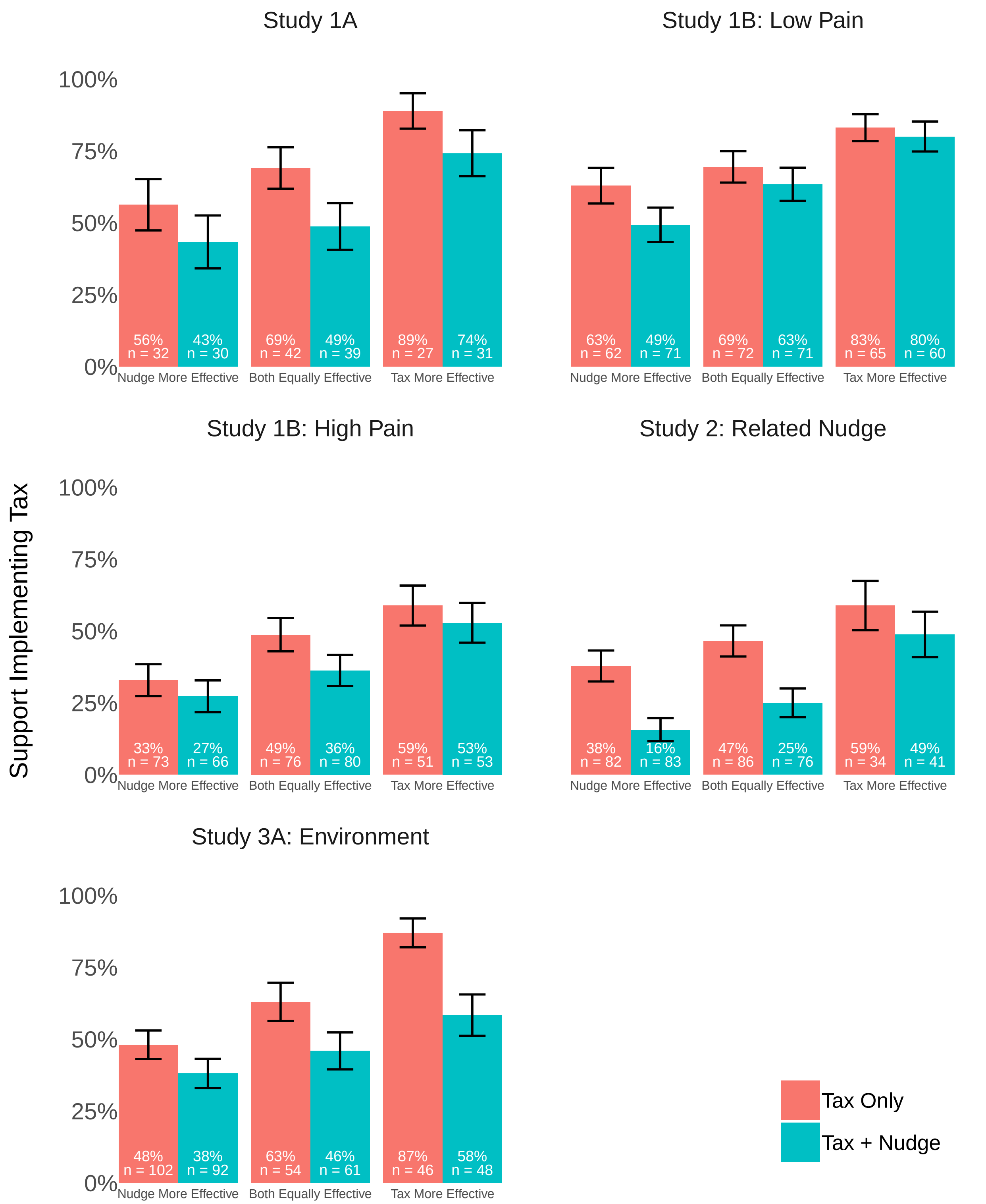


Figure A.9: Support for carbon tax by condition and by whether the tax was perceived to be less, equally, or more effective than the green energy nudge.



multiplicative interaction model, where the moderating variable is the perceived tax effectiveness is subtracted from the perceived nudge effectiveness, we find no statistically significant difference between a linear interaction model and a nonlinear one, $p = 0.38$. In all the cases, more granular binnings also did not yield significant results.

It appears that the introduction of a green energy nudge crowds-out support for a carbon tax even among those who might otherwise be favorable to the tax. Crowd-out is no smaller for those who identify themselves as more liberals, who may generally favor government intervention, or for those who strongly agree that global climate change is occuring. The effect of the experimental treatment is also no smaller for those who think the tax is more effective than the nudge. If anything, those more supportive of implementing a carbon tax see their support diminished the most: those who believe a tax to be ineffective are unlikely to favor its implementation even in the absence of other policies.

A.8 Study S1

Individuals are frequently polled about their attitudes to or approval of various policies. Such responses may in turn influence elected officials, particularly when they worry that a policy proposal may be disliked by their constituents. Could merely being informed about nudges (e.g., through media coverage) undermine support for heavy-handed interventions like taxes?

In this supplementary study, we present participants with three policies aimed at reducing carbon emissions, and ask them to report how effective they believe each to be at reducing emissions and whether they would approve their implementation. The two remaining policies, both nudges, mandate companies to implement defaults that consumers can opt-out of. One of these nudges targets energy suppliers and requires them to default customers into a renewable energy plan; the other targets airlines and requires them to, via a default that can be overridden by the consumer, impose a carbon offset fee on each plane ticket by default. We hypothesize that first learning about a nudge will decrease approval of the carbon tax. We preregistered the study on AsPredicted.org

#4571.

We recruited participants from Amazon Mechanical Turk (N = 601, 42.76% female, mean age 35.03). Within our sample, 26.96% identified as moderately or very conservative and 47.26% had a 4-year college degree or higher. Participants received a fixed payment of 50 cents.

We first elicited beliefs about climate change. Participants were asked about the extent to which they agreed that global average temperatures were increasing and whether human activity was the primary cause of such warming (both on a 5-point scale ranging from "strongly disagree" to "strongly agree"). Participants who strongly disagreed that temperatures are increasing were not asked about human activity as a driver. We then presented three policies to combat climate change and environmental pollution. Two were nudges taken from Sunstein (2016) and one was the carbon tax. One of the nudges requires large electricity providers to automatically enroll consumers in an environmentally friendly energy plan, but gives them the option to opt-out. The other nudge would require airlines to charge flyers a fee to offset their carbon emissions (approximately $10 a ticket), but people could opt-out of the payment if they wanted to. Sunstein (2016) surveyed participants and found that 72% of respondents approved of defaulting customers into green energy plans, but only 36% approved of the default offset on plane tickets. Consequently, we refer to the former policy as a "favorable nudge" and the latter as an "unfavorable nudge." The standard carbon tax policy imposes a tax of $40/ton of carbon on companies based on the emissions created.

For each policy, participants evaluated (on a five-point Likert scale) how favorable they viewed the policy ("Strongly disapprove" to "Strongly approve") and how effective they believed it to be at reducing pollution and carbon emissions ("Not effective at all" to "Extremely effective"). We randomly assigned participants to 6 conditions, varying the order in which they evaluate the policies. Participants in the "Favorable Nudge First" condition first evaluated the green energy nudge, followed by the carbon tax, followed by the airline nudge. Participants in the "Unfavorable Nudge First" condition first evaluated the airline nudge, followed by the carbon tax, followed by the green energy nudge. In two conditions ("Favorable Nudge Passed" and "Unfavorable Nudge

Passed"), we explicitly told participants after evaluating the first nudge that they should assume it had passed, after which they evaluated the carbon tax. Finally, in the last two "Policy First" conditions, participants first evaluate the carbon tax followed by the two nudges in each of the two possible orders. We also collected demographic information on gender, age, ethnicity, education level, and political affiliation.

A.8.1 Results

Most participants in our sample thought that global mean temperature has been increasing over the past 50 years (84.36% somewhat or strongly agreed with this statement). We asked all but those who strongly disagreed with the previous statement (96.67%) whether they believed that human action is primarily responsible for this observed warming: of these, 76.76% somewhat or strongly agreed.

We find that, in line with survey responses from Sunstein (2016), participants viewed the nudge defaulting consumers into green energy plans more favorably than the carbon offset nudge (3.74 vs. 3), which is significantly higher using a paired t-test, ($t(600) = 13.87$, $p < .001$). Participants also believed the energy nudge to be significantly more effective (3.06 vs. 2.11, $t(600) = 19.82$, $p < .001$). Notably, the carbon tax enjoyed the highest approval (3.83) and assessed effectiveness (3.07), though neither is significantly higher than for the favorably-viewed green energy nudge (approval: $t(600) = -1.93$, $p = .054$, and effectiveness: $t(600) = -0.25$, $p = .801$).

Table A.13 presents the results of an ordered logistic regression of approval of the carbon tax on each of the conditions as a predictor variable (Model 1), and the same specification, but including a control for how effective the participant believed the carbon tax would be (Model 2). Not surprisingly, participants who believed the tax to be more effective at reducing pollution also were more supportive of implementing it. When we look at our treatments, however, we find that approval for the tax was significantly lower in each of the four conditions in which the nudge was shown first with the control for effectiveness. We ran a general linear hypothesis analysis to compare the

Table A.13: Ordered logistic regression on the support for the carbon tax as a function of the experimental condition (compared to the baseline in which they see the tax first). Participants view the carbon tax less favorably when a nudge is introduced first, independent of whether they generally hold a favorable or unfavorable view of the nudge.

	Model 1	Model 2
Favorable Nudge First	-0.176	-0.563^{*}
	(0.233)	(0.250)
Favorable Nudge Passed	-0.205	-0.534^{*}
	(0.222)	(0.237)
Unfavorable Nudge First	-0.496^{*}	-0.782^{**}
	(0.235)	(0.245)
Unfavorable Nudge Passed	-0.351	-0.609^{**}
	(0.210)	(0.222)
Tax Effectiveness		1.242^{***}
		(0.084)
Log Likelihood	-829.760	-696.548
Num. obs.	601	601

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

coefficients across our conditions and find no statistically significant differences (all $p > 0.80$). Thus, support for the carbon tax declined when a nudge was introduced first, but this did not depend on whether the nudge was favorable or unfavorable or whether or not participants were told to assume it had passed. We subsequently combine the conditions into a new variable called "Nudge First."

In Table A.14, we show the regression results comparing the four conditions in which the nudge was shown first to the two conditions in which the policy was observed first. In Models 1 and 2, our dependent variable is the approval for the carbon tax. Confirming our previous results, seeing the nudge first indeed decreases support for the tax, controlling for the perceived effectiveness of the tax ($p < 0.001$). This effect also holds when we control for a belief in climate change (splitting participants into those who somewhat or strongly agree with a warming trend and those who do not) and political conservatism (performing a mean split). Note that even after introducing those controls, the coefficient on our experimental conditions remains virtually unchanged.

Could participants be susceptible to a framing effect in the other direction? That is, would support

Table A.14: Ordered logistic regression on the support for the carbon tax collapsed across conditions in which the nudge was shown first. Participants view the carbon tax less favorably when a nudge is introduced first, including with controls (Models 1 and 2). Observing the tax first does not, however, influence approval of either the favorable (Model 3) or unfavorable nudge (Model 4).

	Model 1	Model 2	Model 3	Model 4
Nudge First	-0.619^{***}	-0.621^{***}	0.285	0.023
	(0.172)	(0.173)	(0.158)	(0.167)
Tax Effectiveness	1.242^{***}	1.158^{***}		
	(0.084)	(0.086)		
Offset Nudge Effectiveness			1.122^{***}	
			(0.082)	
Energy Nudge Effectiveness				1.322^{***}
				(0.093)
Climate Change Exists		0.869^{***}	0.677^{**}	1.009^{***}
		(0.218)	(0.208)	(0.221)
Conservative		-0.855^{***}	-0.395^{**}	-0.470^{**}
		(0.166)	(0.153)	(0.166)
Log Likelihood	-697.022	-672.334	-797.445	-663.436
Num. obs.	601	601	601	601

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Coefficients are expected log-odds.

for a nudge decrease when a tax policy is presented first? It might be that our elicitation mechanism (two five-point Likert scales) inherently biases responses and that this is driving our effect. To test this, we investigate whether the order in which participants viewed the tax and the nudge affected approval for the two nudges. Based on our hypothesis, we predict that nudges crowd-out support for taxes, but not vice-versa. In Model 3, we use as our dependent variable support for the (less favorably viewed) offset nudge and, in Model 4, the dependent variable is support for the green energy nudge. In these models, we substitute perceived effectiveness of the respective nudges as a control variable, rather than the effectiveness of the carbon tax. Similar to support for the carbon tax, the perceived effectiveness, belief in climate change, and political liberalism are associated with increased support for the nudge. Importantly, however, whether the nudge was presented first does not determine approval for either nudge, suggesting that knowledge of nudges displaces support for taxes but that the inverse relationship does not hold.

Table B.1: Sample characteristics. Standard deviation in parenthesis.

	CMU (n=542)	NYT (n=216)
Male	61%	47%
Mean Age	48.74	51.81
	(15.25)	(14.03)
Race		
Caucasian	89%	90%
Black	1%	1%
Asian	8%	5%
Other Race	2%	4%
Religion		
Non Religious	35%	52%
Catholic	18%	9%
Protestant	22%	14%
Jewish	12%	9%
Other Religion	13%	16%
Been in ICU	12%	13%
Experiencing Health Problems	7%	10%

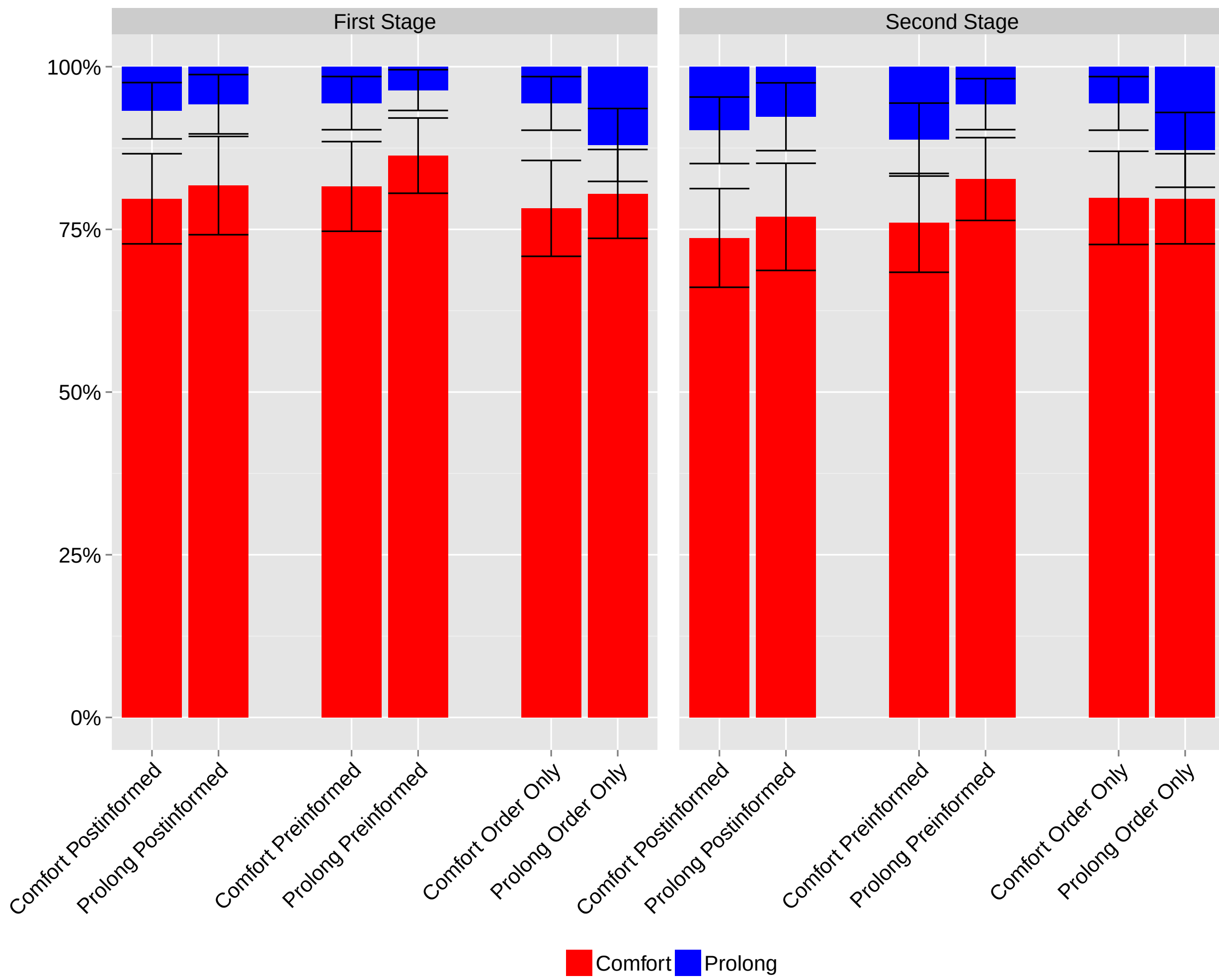
B Appendix: Warning: You Are About to be Nudged

The sample characteristics are shown in Table B.1. Included in this table, as well as in Figures B.1 and B.1. are two conditions we ran and did not report in the main body of the article. The two conditions manipulate the order in which the options are presented (comfort first or prolong first) and do not impose a default. We report them here for sake of completeness, but because our focus is on the effect of defaults, we did not run any analyses using these two conditions.

Participants in this study showed an overwhelming preference for minimizing discomfort. When the question was posed in general terms, 75% of responses reflected this goal (see Figure B.1). By comparison, only 15% of responses reflected a goal of prolonging life. Participants' overall preference for comfort was so fixed that neither default options nor participants' knowledge of them had much impact on their response to the question about their general preference for end-of-life care.

Several main findings are apparent from Figure B.1 and Tables B.2 and B.3. First, a comparison

Figure B.1: The impact of defaults on overall goal for care. The first stage responses include a default (except for the order only conditions); in the second stage, participants answer the same question with no default. Irrespective of condition or default, we see about 75% of participants preferring the comfort option. Error bars are included to indicate 95% confidence intervals.



between the first-stage responses of those defaulting either to comfort or to prolong in the postinformed conditions provides a simple test of conventional default effects. These can be seen in the left-hand panel of Figure B.2 (indicating responses to the first advance directive) comparing the blue and red bars marked "Comfort Postinformed" and "Prolong Postinformed" and in columns 2 and 4 of Table B.2. Significance tests of the difference, presented in the second column of Table B.3, show that five of the five specific items display significant differences at the .10 level, with several significant at more conservative levels. The difference between comfort and prolong for the average of all five items is significant at the .001 level.

Second, a comparison between the blue and red bars marked "Comfort Preinformed" and "Prolong Preinformed" in the left-hand panel of Figure B.2 and between the first and third columns of Table B.2 shows that the default was effective in changing first-stage responses, despite the warning about the default. Although all of the changes are affected by the default in the predicted direction, it can be seen in Table B.3 that the effect is statistically significant only for a single item (dialysis) and for the average of all five items (both at the .01 level). Thus, there is suggestive evidence that preinforming subjects about the default may have weakened but not eliminated its impact.

The third important comparison is between second-stage responses in the postinformed condition, at which point respondents had been informed about the existence of the defaults. The blue and red bars marked "Comfort Postinformed" and "Prolong Postinformed" in the right-hand panel of Figure B.2 and the sixth and eighth columns of Table B.2 show the effect of the default when respondents have been informed of the default after making first-stage choices and are then given the opportunity to revise their choices. The fourth column of Table B.3 shows that the effect of the default is significant at the .05 level or greater for three of the five specific items, and the combination of the five items is significant at the .001 level. The effect of the default, therefore, persisted even when respondents were informed about the default and given an opportunity to reconsider their previous choices.

Figure B.2: The impact of defaults on the average share of choices favoring the comfort or prolong option. In the postinformed conditions, participants are informed between the first and second stages. In the preinformed conditions, they are informed before the first stage. Error bars are included to indicate 95% confidence intervals.

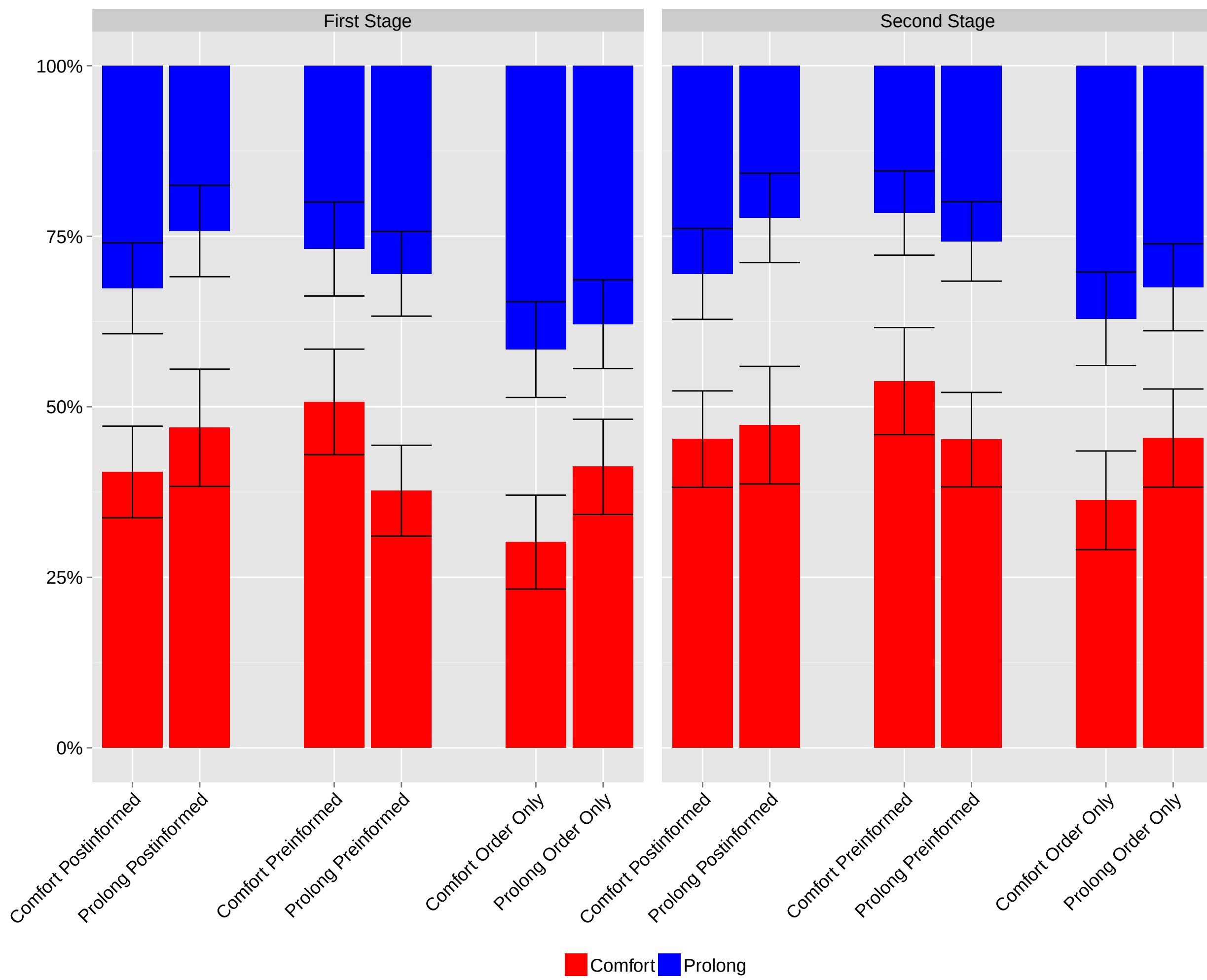


Table B.2: Percentage Choosing Comfort (Top Left of Cells) and Prolong (Bottom Right of Cells) by Stage, Condition, and Item. CPR = cardiopulmonary resuscitation; ICU = intensive care unit.

	Stage 1				Stage 2			
Comfort / Prolong	Comfort preinformed	Comfort postinformed	Prolong preinformed	Prolong postinformed	Comfort preinformed	Comfort postinformed	Prolong preinformed	Prolong postinformed
Overall goal	81.6% / 5.6%	81.7% / 5.8%	80.5% / 12.0%	78.2% / 5.6%	76.0% / 11.2%	76.9% / 7.7%	79.5% / 12.9%	79.8 / 5.6%
Average of 5 specific items	50.7% / 26.9%	46.9% / 24.2%	41.2% / 37.9%	30.2% / 41.6%	53.8% / 21.6%	47.3% / 22.3%	45.4% / 32.5%	36.3% / 37.1%
CPR	44.0% / 35.2%	41.3% / 30.8%	33.1% / 47.4%	23.4% / 53.2%	45.6% / 30.4%	40.4% / 30.8%	39.1% / 42.9%	25.0% / 51.6%
ICU	40.8% / 35.2%	38.5% / 30.8%	33.8% / 40.6%	21.0% / 50.8%	45.6% / 30.4%	40.4% / 26.0%	34.6% / 39.1%	25.8% / 47.6%
Ventilator	59.2% / 19.2%	54.8% / 16.3%	51.1% / 25.6%	37.9% / 29.8%	61.6% / 13.6%	55.7% / 13.5%	55.6% / 21.1%	43.5% / 27.4%
Dialysis	50.4% / 26.4%	47.1% / 26.9%	36.1% / 46.6%	27.4% / 56.8%	53.6% / 20.8%	47.1% / 26.9%	41.4% / 36.1%	37.9% / 36.3%
Feeding tube	59.2% / 18.4%	52.9% / 16.3%	51.9% / 29.3%	41.1% / 27.4%	62.4% / 12.8%	52.9% / 14.4%	56.4% / 23.3%	49.2% / 22.6%

Table B.3: Results of Chi-Square Tests on the Proportion of Comfort, Prolong, and No-Choice Decisions for Participants Who Have Been Defaulted Into the Comfort or Prolong Condition Note. CPR = cardiopulmonary resuscitation; ICU = intensive care unit.

χ^2 Test	Stage 1		Stage 2	
Decision	Preinformed	Postinformed	Preinformed	Postinformed
Overall goal	4.78 ($p < .10$)	0.60 ($p = ns$)	1.99 ($p = ns$)	0.45 ($p = ns$)
5 items combined	18.71 ($p < .001$)	46.42 ($p < .001$)	19.45 ($p < .001$)	30.11 ($p < .001$)
CPR	4.35 ($p = ns$)	12.86 ($p < .01$)	4.45 ($p = ns$)	10.67 ($p < .01$)
ICU	1.40 ($p = ns$)	11.55 ($p < .01$)	3.49 ($p = ns$)	11.65 ($p < .01$)
Ventilator	2.01 ($p = ns$)	8.11 ($p < .05$)	2.50 ($p = ns$)	7.01 ($p < .05$)
Dialysis	11.33 ($p < .01$)	11.94 ($p < .01$)	7.54 ($p < .05$)	2.69 ($p = ns$)
Feeding Tube	4.23 ($p = ns$)	4.79 ($p < .10$)	4.88 ($p < .10$)	2.52 ($p = ns$)

Regression analyses. There were three possible responses for each decision: comfort, prolong, or forgo making a choice (that is, no choice). Because these do not form any natural ordering, we ran a multinomial logistic regression in which the likelihood of choosing prolong or making no choice is compared with that of picking comfort. In the results that follow, each response from an individual was treated as an observation. We corrected for nonindependence among the multiple responses by including random effects at the individual subject level. We also included fixed effects for each of the five items to take into account that the respective options are not equally appealing for all health interventions.

Standard default effects. Table B.4 presents results from multinomial random effects logistic regressions. The first two columns of coefficients in the table examine the stage 1 decisions of individuals in the postinformed condition and compare the choices made by those defaulted into prolong with the decisions made by those defaulted into comfort. This difference is the standard default effect. From the table, it can be seen that the effect of the comfort default is a decrease of 2 in the log-odds ratio for both prolong and no choice relative to the comfort choice. Individuals

Table B.4: Multinomial Logit Regression Showing the Change in Likelihood of Selecting the Comfort Option when the Default Changes from Prolong to Comfort. Note. The dependent variable is the combined answers to all treatment decisions. Included are controls for population (Carnegie Mellon University and New York Times), gender, age, age squared, ethnicity, and religiosity. The regression also includes fixed effects for each of the five questions and random effects at the individual level. *p < .05. **p < .01. ***p < .001.

	Stage 1 postinformed		Stage 2 postinformed		Stage 1 preinformed	
	Prolong	No choice	Prolong	No choice	Prolong	No choice
Comfort default	-2.032**	-2.050*	-1.100***	0.333	-1.553**	-0.389
	(0.708)	(0.922)	(0.317)	(0.520)	(0.488)	(0.371)
Marginal effect	0.361**		0.175**		0.329**	
(comfort=0)	(0.112)		(0.068)		(0.111)	

defaulted into the comfort option are significantly less likely to choose prolong (p < .01) or no choice (p < .05) relative to the comfort choice.

The second row of the table (labeled Marginal Effect) shows the predicted percentage point change in choice of the comfort option resulting from changing the default from prolong to comfort. The number 0.361 indicates that shifting an individual from the prolong to the comfort default increases that individual's probability of choosing the comfort option by 36 percentage points (p < .01). The p value here denotes the significance test for the null hypothesis that shifting them into the comfort default has no impact.

Postinformed. After being informed of the defaults, a substantial fraction of respondents (15.4% for those defaulted to comfort and 20.1% of those defaulted to prolong) picked a different number of comfort choices. Among those defaulted to comfort, however, the number changing in each direction approximately canceled out, so there were no significant net changes. In contrast, those defaulted to prolong showed a robust (consistent across all five items) propensity to shift toward the comfort option, consistent with a reactance effect. The change was greatest for dialysis (with a net change of 11% toward the comfort option) and smallest for cardiopulmonary resuscitation (with only 2% shifting toward the comfort option).

To examine the overall effect of defaulting choices for people, informing them that certain answers have been set as defaults, and then letting them choose again (this time with no defaults), we ran regression analyses comparing the second phase of the comfort and prolong postawareness default conditions (see the middle columns of Table B.4). The coefficient on comfort in the prolong column is again significantly negative ($p < .001$), showing that participants in the comfort condition are significantly less likely to choose the prolong option. Despite postinforming respondents of the defaults and allowing them to revise their responses, we found that their second responses were still affected by the original default, although the effect was quantitatively smaller (17.5 percentage points; see the second row of Table B.4) than reported for the first round. Their likelihood of choosing to let an agent make the decision for them relative to choosing comfort is not significantly different.

Preinformed. The last two columns of Table B.4 report the effect of defaults in the preinformed treatment. The likelihood of choosing prolong compared with comfort is significantly lower for those in the comfort default ($p < .01$). The magnitude of the log-odds change is smaller than in the postinformed condition, and the shift from not making a choice into comfort is no longer significant. However, the marginal effects in the bottom row of Table B.4 show that the comfort default increases the probability of choosing the comfort option by 33 percentage points, which is about the same as in the first-stage responses to the postinform conditions, in which respondents had not been alerted to the defaults. According to this analysis, preinforming people of defaults had, at most, a small impact on their effectiveness.

As a more direct test of whether preinforming respondents affected the impact of the defaults, we pooled both default conditions (comfort and prolong) and defined a variable that was equal to 1 if the respondent, in the first phase, chose an option other than the one to which they had been defaulted. We then regressed this variable through a series of several (related) specifications: initially using only a binary indicator of whether they had been preinformed, then adding all control variables, and finally including a binary indicator for the comfort default, plus the interaction term between comfort and preinformed. In none of these specifications did either the preinform variable or the interaction term approach significance, suggesting that preinforming respondents about the default does not diminish their tendency to stick with the default.

Analysis of individual items. Table B.5 presents multinomial logistic regressions for each individual option. As before, we compare the changes in log-odds of choosing the prolong and no-choice options relative to the comfort baseline. Each regression includes terms for the comfort default condition (estimating the effect of the default for someone who is postinformed), the preinformed condition (estimating the effect of preinforming someone in the prolong condition), and their interaction (allowing us to calculate the effect of the default for someone who is preinformed).

From the table, it can be seen that the comfort default for those who have not been preinformed significantly reduces the likelihood of choosing the prolong option, relative to the comfort option, in three of the five items ($p < .01$ for CPR and ICU and $p < .05$ for ventilator use). The fourth row of the table shows that being defaulted into comfort increases the overall probability of choosing comfort in the CPR and ICU decisions by approximately 15 percentage points. The marginal effects for the remaining choices are substantially smaller and nonsignificant. The default does not affect significantly the likelihood of making no choice, compared with picking the comfort option, on any of the individual items.

Table B.5: Multinomial Logistic Regression on the Choice when Filling out the Advance Directive for the Second Time (with No Default). Note. The baseline is the comfort choice. The marginal effects show the change in likelihood of selecting the comfort option when the default changes from prolong to comfort for someone who is at the mean of our sample and not preinformed, as well as the effect of preinforming someone who is at the mean of the sample and defaulted into the comfort option. Included are controls for population (Carnegie Mellon University and New York Times), gender, age, age squared, ethnicity, and religiosity. CPR = cardiopulmonary resuscitation; ICU = intensive care unit. *p < .05. **p < .01. ***p < .001.

	CPR		ICU		Ventilator		Dialysis		Feeding tube	
	Prolong	No choice	Prolong	No choice	Prolong	No choice	Prolong	No choice	Prolong	No choice
Comfort	-1.084**	-0.339	-1.111**	-0.260	-0.870*	-0.166	-0.431	-0.209	-0.394	0.146
	(0.351)	(0.361)	(0.361)	(0.348)	(0.398)	(0.317)	(0.333)	(0.340)	(0.400)	(0.314)
Preinformed	-0.675*	-0.803*	-0.455	-0.401	-0.354	-0.519	-0.015	-0.275	0.058	-0.494
	(0.322)	(0.367)	(0.329)	(0.347)	(0.338)	(0.313)	(0.306)	(0.331)	(0.342)	(0.319)
Preinformed ×	0.597	0.484	0.546	-0.085	0.319	0.179	-0.363	0.101	-0.312	0.009
Comfort	(0.466)	(0.494)	(0.478)	(0.476)	(0.538)	(0.441)	(0.459)	(0.465)	(0.541)	(0.444)
Marginal effect	0.146*		0.163*		0.096		0.075		0.007	
(comfort = 1)	(0.060)		(0.060)		(0.065)		(0.065)		(0.065)	
Marginal effect	0.044		0.049		0.056		0.065		0.095	
(preinformed = 0)	(0.063)		(0.063)		(0.065)		(0.065)		(0.064)	

The second and fifth rows of Table B.5 address the question we are most interested in: whether preinforming someone of the comfort default decreases their probability of choosing the comfort option. If so, preinforming could decrease the intervention's effectiveness. Preinforming subjects in the prolong default does decrease the number of prolong and no-choice decisions but does not affect the strength of the default effect in the comfort condition (the sum of the preinformed coefficient and the interaction coefficient).

Advance directive preferences—Compulsory rules or mere guidelines? Finally, participants indicated, after the second phase, whether their agent "must follow these instructions" (binding) or whether these instructions should be treated "only [as] guidelines" (nonbinding). We examined the association between the second-phase responses and the binding/nonbinding designation, as shown in Figure B.3. Respondents who considered the directive to be binding were more likely to choose the comfort option, and separate analyses revealed that the transpose is also true: Decisions favoring comfort were more likely to be designated as binding than were decisions favoring prolongation of life.

Figure B.3: Share of decisions by level of commitment. The figure shows the proportion of comfort and prolong choices for people who wanted their decisions to be binding and not binding on their health care provider, respectively. Error bars are included to indicate 95% confidence intervals.