Management Science

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

Information Nudges and Self-Control

Thomas Mariotti, Nikolaus Schweizer, Nora Szech, Jonas von Wangenheim

To cite this article:
Thomas Mariotti, Nikolaus Schweizer, Nora Szech, Jonas von Wangenheim (2022) Information Nudges and Self-Control. Management Science

Published online in Articles in Advance 26 May 2022

. https://doi.org/10.1287/mnsc.2022.4428

http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–16
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Information Nudges and Self-Control

Thomas Mariotti,[a,b,c] Nikolaus Schweizer,[d] Nora Szech,[c,e,f,*] Jonas von Wangenheim[g]

[a] Toulouse School of Economics, Centre National de la Recherche Scientifique, University of Toulouse Capitole, Toulouse 31042, France; [b] Centre for Economic Policy Research, London EC1V 0DX, United Kingdom; [c] Center for Economic Studies, Munich 81679, Germany; [d] Department of Econometrics and Operations Research, Tilburg University, 5037 AB Tilburg, Netherlands; [e] Karlsruhe Institute of Technology, ECON Institute, Karlsruhe 76049, Germany; [f] Berlin Social Science Center, 10785 Berlin, Germany; [g] Institute for Microeconomics, University of Bonn, Bonn 53113, Germany
*Corresponding author

Contact: thomas.mariotti@tse-fr.eu, https://orcid.org/0000-0002-0525-8743 (TM); n.f.f.schweizer@uvt.nl, https://orcid.org/0000-0002-5807-7321 (NiS); nora.szech@kit.edu, https://orcid.org/0000-0002-6674-4569 (NoS); jwangenheim@uni-bonn.de, https://orcid.org/0000-0002-2830-9961 (JvW)

Received: February 24, 2020
Revised: June 26, 2021; March 10, 2022
Accepted: March 26, 2022
Published Online in Articles in Advance: May 26, 2022

https://doi.org/10.1287/mnsc.2022.4428

Abstract. We study the optimal design of information nudges directed to present-biased consumers who make consumption decisions over time without exact prior knowledge of their long-term consequences. For any distribution of risks, there exists a consumer-optimal information nudge that is of cutoff type, recommending abstinence if the risk is high enough. Depending on the distribution of risks, more or fewer consumers have to be sacrificed, as they cannot be credibly warned even though they would like to be. Under a stronger present bias, the target group receiving a credible warning to abstain must be tightened, but this need not increase the probability of harmful consumption. If some consumers have a stronger present bias than others, traffic-light nudges turn out to be optimal and, when subgroups of consumers differ sufficiently, the optimal traffic-light nudge is also subgroup optimal. We finally compare the consumer-optimal nudge with those that a health authority or a lobbyist would favor.

History: Accepted by Manel Baucells, behavioral economics and decision analysis.

Funding: This research benefited from the financial support of the Agence Nationale de la Recherche [Programme d'Investissements d'Avenir Grant ANR-17-EURE-0010], the German Research Foundation [Grants Collaborative Research Center Transregio 190 and 224 Project B03], and the research foundation Toulouse School of Economics-Partnership.

Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2022.4428.

Keywords: nudges • present bias • self-control • information design

1. Introduction

There has been a remarkable variety across space and time in the attempts to alleviate the consumption of potentially harmful goods. A particularly drastic policy is to prohibit those goods altogether. This was done in the United States in the 1920s with regard to alcohol. However, Prohibition did not prevent illegal consumption: data suggest that although consumption first declined during Prohibition, it increased again after a few years, once the illegal market had adapted; consumption remained stable after Prohibition ended (Miron and Zwiebel 1991). On top of being illiberal and leading to the criminalization of many people, this extreme measure only achieved moderate results regarding drinking behavior (Hall 2010). A similar case has been more recently made against drug prohibition (Miron and Zwiebel 1995). The reason might be that prohibition does not credibly convey information about the actual hazards of consumption. Nowadays, a more liberal and more informative approach is to use information nudges. For example, in many countries, cigarette packages now come with graphic information and text messages about the potential consequences of smoking. Consumers take those warnings as sources of information and react to such labels, at least to some extent.[1]

However, empirical research also documents that consumers do not always feel properly addressed. In a study with adolescents, McCool et al. (2012, p. 1271) report that many participants questioned whether the graphic labels "portrayed an *authentic* representation of the harm caused by smoking." Indeed, the majority perceived such labels as "showing the worst case scenario" because, for example, "of course no-one's going to let their foot get that bad." A targeted and

more credible information nudge may have more potential. For example, warnings against drinking during pregnancy seem to have a significant impact on those concerned (Hankin et al. 1993). Yet little is known about the optimal design of information nudges, and what target groups to address best. This paper aims at filling this gap.

Our formal analysis relies on three ingredients: present bias, incomplete information, and Bayesian updating. Let us examine each of these ingredients in turn.

Present Bias. In our model, a present-biased decision maker (DM) puts a disproportionate utility weight on the current period compared with all later periods (Ainslie 1975, 1992; Thaler 1981; Loewenstein and Prelec 1992; Yoon 2020). The DM has to make consumption decisions over time that may have harmful consequences in the future. In this context, the DM's preferred course of action may look as follows: Cheat today, but abstain from tomorrow on. Under no commitment, however, this course of action is not feasible: once tomorrow is reached, the same logic applies, so that cheating "today" combined with abstaining from "tomorrow" on looks most appealing—again! As a consequence, every day becomes a cheating day, and consumption never comes to an end. The DM may be aware that quitting once and for all would be a smarter choice than engaging in harmful consumption forever. Yet again this choice may not be feasible under no commitment, with potentially dreadful consequences.

Incomplete Information. The DM initially has incomplete information about the harmful consequences of consumption. This can be because their likelihood hinges on the DM's individual risk, which the DM need not know with precision. For instance, there may be heterogeneity in risks across individuals; then, although risk statistics may be available at the aggregate population level, the DM's exact position in the distribution of risks may be unknown, because it depends on a variety of factors the DM lacks the expertise to assess and combine. Alternatively, one could think of a population of DMs facing an aggregate risk of unknown magnitude. In both interpretations, we assume that the distribution of risks is common knowledge.

Bayesian Updating. In this context, information nudges can help the DM overcome the present bias by modifying the DM's beliefs. We assume that such a nudge has to be credible, and thus cannot systematically overstate the harmful consequences of consumption. Thus, it can be thought as a mechanism designed to send messages—specifically, incentive-compatible recommendations to consume or abstain—to the DM conditional on the true value of the risk. The DM, once exposed to new information through the nudge, updates prior beliefs in a Bayesian way and acts accordingly. This generates a tradeoff between the credibility of the nudge and its efficiency at deterring consumption whenever it is undesirable.

We characterize the optimal information nudge from the DM's perspective prior to taking any consumption decision. We show that there always exists a consumer-optimal information nudge of cutoff type, in which DMs learn that the risk they are facing is either low or high, depending on whether the risk lies below or above a certain cutoff. The intuition is that cutoff mechanisms have good efficiency properties, because consumption only takes place when the risk is low enough, and that they also have good incentive properties, because abstention is incentive compatible only when the risk is perceived to be high enough. When individual risks are heterogeneous, these signals can be interpreted as warnings against consumption for high-risk individuals within the target group of the information nudge. When the risk is an aggregate one, credible information is conveyed to the whole population. In either case, finding the optimal information nudge is easy, in that it requires pinning down one single parameter. What makes this task challenging is that it requires precise knowledge of the DM's present bias.

The reason why the optimal cutoff structure of the nudge outperforms full transparency is that the credibility of the warning that pools risks above the cutoff enables more DMs to find the strength to abstain from consuming once they have learned that their risk is relatively high, whereas they would have engaged in harmful consumption under full information. This contrasts with a DM with no present bias, for whom full transparency would be optimal. Tightening the target group of the nudge enables one to credibly communicate more drastic information, thereby more effectively counteracting impulses from the DM's present bias. Indeed, such tightening may explain why warnings against alcohol work best when they are targeted at the most vulnerable groups, such as pregnant women. Of course, in practice, many other individuals should better abstain (see, e.g., Gutjahr et al. 2001). Yet our analysis suggests that it may be optimal to warn only high-risk DMs in order to deter at least them successfully, sacrificing DMs with lower but still significant risk who end up trapped in harmful consumption. Key to this logic is that the optimal information structure is coarse: it is more efficient to shield the maximum mass of high-risk DMs away from consumption by issuing a straight recommendation to abstain, rather than to issue mixed messages that would only partially protect inframarginal DMs with relatively lower risk.

We provide two types of comparative statics results for the consumer-optimal information nudge. We first

prove that a shift of the risk distribution toward higher levels of risk leads to a strictly lower probability of consumption. This reflects two complementary effects. First, such a shift makes abstinence more desirable; second, in line with the aforementioned properties of cutoff mechanisms, it makes it easier to sustain abstinence in an incentive-compatible way. We next investigate the impact of a shift in the DM's present bias on the probability that harmful consumption takes place.[2] We show that, if the distribution of risks satisfies the monotone-hazard-rate property, harmful consumption takes place with strictly positive probability under the optimal information nudge if and only if the DM's present bias is severe enough. We also provide a sufficient condition for the probability of harmful consumption to be decreasing in the DM's degree of self-control.

Whereas the previous results rely on a fixed and known present bias, there is in fact significant heterogeneity in self-control across individuals (Mischel 2014, Sutter et al. 2013). For example, consumers with high self-control differ from consumers with low self-control when it comes to food choice, as has been shown in a study on the potential impact of product labeling on health (Koenigstorfer et al. 2014). This leads us to analyze the more realistic scenario in which DMs may have high or low self-control, a characteristic that is their private information. We suppose that a single information nudge has to be designed for an entire population of DMs. This case is empirically relevant for tobacco, alcohol, or food warnings, which are often printed on the items consumers can choose from. DMs with both low and with high self-control can then be optimally informed via the same information nudge, which turns out to be a green-yellow-red traffic-light nudge. Whereas the strongest, red warning is drastic enough to make DMs abstain regardless of their degree of self-control, the intermediate yellow warning convinces at least DMs with high self-control to abstain.

Our results are threefold. First, when the two types of DMs have very different degrees of self-control, their individually optimally information nudges can be combined into a single traffic-light nudge without affecting incentives; hence the two types exert no externality on each other. This traffic-light nudge is monotone in that it has a two-cutoff structure: a green light is issued for low levels of risk, a yellow warning for intermediate levels of risk, and a red warning for high levels of risk, with the same cutoffs as under the individually optimal nudges. Next, when the two types of DMs become more alike in terms of self-control, a monotone traffic-light nudge remains optimal, but the corresponding cutoffs have to be modified to preserve incentive-compatibility, which in particular requires that DMs with high self-control abstain when a yellow warning is issued; hence the two types exert an externality on each other, and it becomes necessary, depending on their relative shares in the population, to sacrifice one type to the benefit of the other. This discrimination property may be a reason why traffic-light nudges, which are intuitively perceived as monotone, are one if not the most frequently used nonnumerical information structures, in addition to their potential saliency.[3] Finally, when the two types of DMs are similar in terms of self-control, the monotonicity of the optimal traffic-light nudge may be lost and a three-cutoff mechanism may become optimal, whereby a yellow warning is issued both for intermediate and for extreme levels of risk. In that case, the intuitive content of traffic-light nudges is more questionable.

Our analysis in most of the paper takes a liberal perspective, focusing on the information nudge that maximizes the DM's expected utility prior to taking any consumption decision. However, it is also interesting to derive the information nudges that are optimal from other perspectives. Examples include a health authority that wishes to minimize the probability of consumption, a lobbyist who wishes to maximize the probability of consumption, or a social planner who wishes to maximize a weighted sum of the DM's utilities at different dates. The cutoff structure of optimal information nudges carries over to these alternative scenarios; yet, of course, the cutoffs are chosen differently. For instance, whereas a health authority prefers to make as many consumers as possible shy away from harmful consumption, a lobbyist prefers to lower willpower in as many consumers as possible by convincing them that the risk is not that high, so that many consumers who would favor an information nudge that helped them abstaining are instead trapped in harmful consumption. A policymaker unaware of consumers' self-control problems may even misinterpret the information structure implemented by a lobbyist as health concerned, when, in fact, the lobbyist deliberately chose the target group of the warning label to minimize the deterrence effect of the nudge. Finally, from a liberal perspective, it would be ideal to choose the cutoff in the consumer-optimal information nudge so that consumption is recommended if and only if it involves no harm. Yet, as we have seen, this mechanism is incentive compatible if and only if the DM's present bias is low enough. In all other cases, harmful consumption takes place with strictly positive probability, and the consumer-optimal information nudge coincides with the one a health authority would favor.

Related Literature. Popularized by Thaler and Sunstein (2008), the literature on nudging is growing fast and into multiple directions. Whereas contributions such as Benkert and Netzer (2018) focus on nudges that influence the framing of decision problems, our focus is on nudges that provide an optimized release of information, so

called information nudges. Such nudges, in the form of warning signals or labels, have received much attention over the years, in particular in the marketing literature (see Argo and Main 2004 for an overview). In general, warning labels seem to be effective. Mazis et al. (1991) find that warning labels on alcohol beverages indeed increase the perception of the risk associated with drinking. In line with our modeling approach, Stewart and Martin (1994, p. 1) conclude that "warnings inform rather than persuade consumers." Kaskutas (1993) finds high support in the population for the use of alcohol warning labels after their introduction in the United States. On a political level, research on nudging has informed policymaking in various countries, such as in the United States, United Kingdom, Australia, Germany, and Japan. Also the Organisation for Economic Co-operation and Development, the United Nations, and the World Bank have set up nudging units.

Avoidance and deliberate pooling of information have been major strands of recent research in various literatures. In a classical rational model, DMs would always prefer to be as well-informed as possible. In this spirit, some contributions, such as Aprahamian et al. (2019), view information costs and efficiency considerations as a rationale for generating and providing coarse information in the medical testing of groups. By contrast, studies such as Zimmermann (2015) or Ganguly and Tasoff (2017) emphasize psychological motives such as anticipatory utility as driving a preference for, for example, delayed or clumped information provision. Golman et al. (2021) and Ho et al. (2020) discuss the modeling and measurement of such psychological preferences for information avoidance.

In our model, the nonstandard preference for information arises due to self-control problems rather than anticipation or motivated beliefs. In this regard, our paper is most closely related to Carrillo and Mariotti (2000) and Bénabou and Tirole (2002). Specifically, we take the basic model of Carrillo and Mariotti (2000) as our starting point. However, instead of focusing on information-acquisition or information-storing processes, we characterize optimal information structures for a present-biased DM using the Bayesian-persuasion approach. Over the last decade, this framework, introduced by Brocas and Carrillo (2007), Rayo and Segal (2010), and Kamenica and Gentzkow (2011), has proved very influential, first in economics and then also in other behavioral sciences, including management.

For instance, Alizamir et al. (2020) analyze reputation concerns of an agency that may issue warnings about recurring harmful events to induce an uninformed stakeholder to take preemptive actions. Lingenbrink and Iyer (2019) study optimal information management in queuing applications. Szydlowski (2021) studies the optimal cash-flow disclosure policy of an entrepreneur who jointly chooses disclosure and financing policies. Papanastasiou et al. (2018) analyze optimal information disclosure on online platforms to level out efficiency and incentives for exploration. Finally, in Drakopoulos et al. (2021), a seller uses Bayesian persuasion to signal product availability in order to manage revenues.

Methodologically, our baseline model is most closely related to the setting of linear, type-dependent sender preferences in Kolotilin (2015), which we follow to derive the consumer-optimal information nudge. Our own analysis really starts with the comparative statics of this nudge, notably with respect to the DM's present bias, and our main methodological contribution to the Bayesian-persuasion literature lies in the analysis of optimal traffic-light nudges under heterogeneous present bias.

What sets our paper apart from most of the Bayesian-persuasion literature is that our focus is on frictions in information demand arising from intrapersonal, psychological conflicts rather than from sender-receiver conflicts of interest. Our paper thereby contributes to a small but growing literature on the optimal disclosure of information to agents with psychological preferences. Lipnowski and Mathevet (2018) show that tempted agents in the sense of Gul and Pesendorfer (2001) does not want to know what they are missing, and thus that an optimal disclosure mechanism should limit agents' information about the value of their preferred choice, so as to reduce the cost of self-control. Schweizer and Szech (2018) study the optimal revelation of life-changing information, such as that provided by a medical test, to a patient with anticipatory utility. Habibi (2020) studies feedback mechanisms when a benevolent principal motivates an agent with present-biased preferences to exert unobservable effort, thereby providing a moral-hazard counterpart to our analysis. Related to the logic in our paper, Gao et al. (2021) argue that medical tests should have a sufficiently high specificity, even at the cost of lower sensitivity, because only then is a positive result sufficiently alarming for subjects to conduct follow-up checks. Similar to our analysis, they find cutoff mechanisms to be optimal.

2. The Model

As in Carrillo and Mariotti (2000), our model features a present-biased DM (he) who has to make consumption decisions over time under no commitment. Consumption is enjoyable in the short term but possibly harmful in the long term. The novelty of the model is that the DM's information about the riskiness of consumption is optimized by a mechanism designer (she).

2.1. Actions and Payoffs

The DM lives at dates $\tau = 0, 1, 2, 3$. At dates $\tau = 0, 1$, the DM can consume, $x_\tau = 1$, or abstain, $x_\tau = 0$.

Consuming at any date τ increases the DM's current utility by 1 but comes with probability θ at a cost C, incurred at date $\tau+2$. As in Phelps and Pollak (1968) and Laibson (1997), the DM has a quasi-hyperbolic discount function with parameters β and δ. That is, the utility indices of his date-0 and date-1 selves are

$$U_0(x_0,x_1,\theta) \equiv x_0 + \beta\delta x_1 - \beta\delta^2\theta C x_0 - \beta\delta^3\theta C x_1, \quad (1)$$

$$U_1(x_0,x_1,\theta) \equiv x_1 - \beta\delta\theta C x_0 - \beta\delta^2\theta C x_1, \quad (2)$$

where $\delta \in (0,1]$ is the per-period discount factor, and $\beta \in (0,1)$ is the time-inconsistency parameter, which is inversely related to present bias. As $\beta < 1$, the DM at date 1 puts, relatively to his utility from consuming, less weight on the harm his consuming might cause at date 3 than he does at date 0. We assume that $\beta\delta^2 C > 1$, so that the DM would always abstain if he believed that the cost C were incurred with probability 1 upon consuming.

2.2. Information and Strategies

The DM's prior beliefs about θ are represented by a distribution $\mathbf{P}$ with support $\Theta \equiv [\underline{\theta}, \overline{\theta}]$ and cumulative distribution function (cdf) F, which admits a continuous density f that is strictly positive, except possibly at $\underline{\theta}$ and $\overline{\theta}$. Thus F is strictly increasing, with well-defined quantiles.[4]

Before the DM's first consumption decision at date 0, the DM is exposed to additional information about θ. This information is distilled by a mechanism designer who knows the value of θ and can commit to a persuasion mechanism issuing messages conditional on that value. The DM then updates his beliefs about θ in a Bayesian way.

As in Strotz (1956), the DM cannot commit to a course of action contingent on his beliefs. This restriction is binding, because the preferences induced by (1)–(2) along with these beliefs are time-inconsistent when $\beta < 1$. As in Peleg and Yaari (1973), the date-0 and date-1 selves of the DM act as independent decision units. The DM is sophisticated, so that his behavior is described by a subgame-perfect equilibrium of the resulting intrapersonal game.

We throughout assume that the DM and the mechanism designer have common prior beliefs about θ. This is an important assumption; otherwise, the optimal persuasion mechanism may take a different form. Moreover, in most of our analysis, we take a liberal perspective; that is, we assume that the mechanism designer is benevolent, in the sense that her interests are aligned with those of the DM at date 0. Alternative objective functions for the mechanism designer are considered in Section 6.

2.3. Applications

Our model applies to a wide range of situations in which a mechanism designer can determine how much information to reveal to a DM regarding the riskiness of consumption. The designer can pool information by issuing a coarse signal. Yet she needs to stick to the truth: that is, she cannot fool Bayesian DMs by systematically lying to them.

Depending on the application, the riskiness may be a characteristic of the product the DM can consume, a characteristic of the DM himself, or a combination of the two. In the first case, information structures are typically identical for a whole population. Think, for example, of information nudges on food and beverages in a supermarket, indicating how healthy a specific choice would be.[5] When the information nudge is printed on the item itself, the mechanism designer decides whether to disclose the riskiness of a product, or to pool information about different products. For example, the designer could decide whether a snack is labeled as a healthy, green-label item or as an unhealthy, red-label item; more detailed information can be provided by a traffic-light nudge. In the second case, the mechanism designer may be able to individually address different consumers, and thereby make use of more personalized signals. An example is information nudging in a supermarket via smart glasses or smartphones. Another case in point is medical advice: for instance, a doctor or a medical agency may have superior information about a patient's riskiness and optimize the way it is communicated to the patient in order to influence the patient's behavior.[6] In the latter case, the riskiness is an individual characteristic of the patient. The doctor can choose to disclose it to the patient perfectly, but can also only disclose that the patient belongs to a group of smaller or larger riskiness. From now on, and bearing in mind these two interpretations of the model, we shall uniformly refer to θ as the DM's *type*.

2.4. The Intrapersonal Game

As a preliminary step, we focus on the intrapersonal game played by the DM's date-0 and date-1 selves following the issue of some message by the mechanism designer. Owing to the binary character of consumption decisions and to the linearity in θ of the date-0 and date-1 selves' utilities, equilibrium behavior in this intrapersonal game only depends on the DM's mean posterior belief $\hat{\boldsymbol{\theta}}$ about θ following this message. Our first result directly follows from (1)–(2).

Lemma 1. *Let*

$$t^a \equiv \frac{1}{\beta\delta^2 C} \in (0,1). \quad (3)$$

If $\hat{\boldsymbol{\theta}} \neq t^a$, then the intrapersonal game has a unique subgame-perfect equilibrium, in which the DM's date-0

and date-1 selves both consume if $\hat{\theta} < t^a$ and both abstain if $\hat{\theta} > t^a$. If $\hat{\theta} = t^a$, then the DM is indifferent between consuming and abstaining at both dates, and any of these behaviors is consistent with a subgame-perfect equilibrium of the intrapersonal game.

Observe from (1) that, if $\beta t^a < \hat{\theta} < t^a$, then the DM at date 0 would be strictly better off consuming at date 0 and abstaining at date 1. However, there is no way the DM can reach this outcome under no commitment. Figure 1 illustrates the DM's date-0 equilibrium payoff correspondence, as well as the date-0 payoffs from consuming at both dates, consuming at date 0 and abstaining at date 1, and abstaining at both dates. Notice that there is a discontinuity in the DM's date-0 equilibrium payoff at $\hat{\theta} = t^a$. Indeed, letting

$$t^h \equiv \frac{1+\beta\delta}{1+\delta} t^a \in (0, t^a), \tag{4}$$

if $t^h < \hat{\theta} < t^a$, then the DM at date 0 would be strictly better off abstaining at both dates than consuming at both dates, and the more so, the closer $\hat{\theta}$ is to t^a. Yet, under no commitment, the DM cannot help doing so; we then say that harmful consumption takes place in equilibrium. The resulting discontinuity in the DM's date-0 equilibrium payoff arises from the DM's present bias: in the limiting case $\beta = 1$, the gap between t^h and t^a vanishes, and the DM's date-0 equilibrium payoff is continuous in $\hat{\theta}$; specifically, it is convex in $\hat{\theta}$, reflecting that the value of information for a time-consistent DM is always nonnegative (Blackwell 1953).

3. Optimal Information Disclosure

If the DM had no present bias or could commit to a course of action, full information at date 0 would be optimal from his perspective. As shown by Carrillo and Mariotti (2000), however, this is no longer the case if the DM suffers from a self-control problem. Because full transparency can destroy beneficial beliefs that help overcome temptation, the value of becoming perfectly informed relative to staying completely ignorant about the risk can be negative from the perspective of the DM at date 0. However, this comparison is extreme, and does not shed light on the date-0 optimal information structure. We now tackle this issue, building on the Bayesian-persuasion literature initiated by Brocas and Carrillo (2007), Rayo and Segal (2010), and Kamenica and Gentzkow (2011). We assume throughout this section that the mechanism designer is benevolent, and thus maximizes self-0's expected utility. We also assume that, in case of indifference, that is, when $\hat{\theta} = t^a$, the DM chooses to abstain, which is his and the mechanism designer's preferred course of action.

3.1. Persuasion Mechanisms

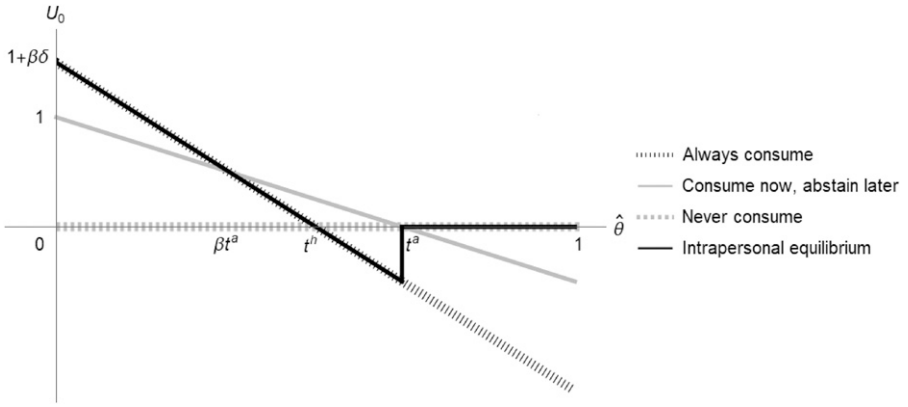
According to the revelation principle, there is no loss of generality in focusing on measurable direct mechanisms $\pi : \Theta \to [0,1]$ that associate to each $\theta \in \Theta$ a probability of issuing an incentive-compatible recommendation to consume at dates 0 and 1. Issuing messages only at date 0 involves no loss of generality. Indeed, if the mechanism designer could commit to different messages at dates 0 and 1, then certainly the DM's date-1 self would take both messages into account for his consumption decision. Because less information at date 0 can only hurt the date-0 self and, hence, the mechanism designer, it is always optimal for the latter to provide the date-0 self with any information that she provides the date-1 self with.

We shall first formulate the relevant incentive constraints and the mechanism designer's optimization problem, and then characterize the optimal information nudge. To make the problem interesting, we assume that the support Θ is sufficiently spread out.

Assumption 1. $\underline{\theta} < t^h < t^a < \overline{\theta}$.

Let π be a mechanism that sends both recommendations to consume and to abstain with strictly positive

Figure 1. Expected Utilities in Different Consumption Scenarios for Varying $\hat{\theta}$



probability, so that applying Bayes' rule is straightforward. By Lemma 1, complying with the recommendation to consume is consistent with equilibrium if and only if

$$\mathbf{E}^{\pi}[\theta|\theta \text{ is recommended to consume}] = \frac{\mathbf{E}[\theta\pi(\theta)]}{\mathbf{E}[\pi(\theta)]} \leq t^a, \tag{5}$$

and complying with the recommendation to abstain is consistent with equilibrium if and only if

$$\mathbf{E}^{\pi}[\theta|\theta \text{ is recommended to abstain}] = \frac{\mathbf{E}[\theta[1-\pi(\theta)]]}{\mathbf{E}[1-\pi(\theta)]} \geq t^a. \tag{6}$$

A mechanism π is incentive compatible (IC) if it satisfies Constraints (5)–(6); if $\pi = 0$ or $\pi = 1$ **P**-almost surely, then, by convention, the undefined constraint is treated as emptily satisfied. Given the Expression (1) for self-0's utility, the mechanism designer's problem becomes

$$\max\{t^h\mathbf{E}[\pi(\theta)] - \mathbf{E}[\theta\pi(\theta)] : \pi \text{ is IC}\}. \tag{7}$$

The objective function in (7) and Constraints (5)–(6) are all affine in π. As a result, deriving the optimal IC mechanism lies within the domain of earlier results in the Bayesian-persuasion literature. This quickly leads us to Proposition 1, which specializes Kolotilin (2015) to our setup. In the appendix, we provide a self-contained derivation of this result and its extension to the more general objective functions presented in Section 6.

It turns out that we can restrict attention to the class of cutoff mechanisms π_t, $t \in \Theta$. The cutoff mechanism π_t recommends to consume if θ is below the cutoff value t,

$$\pi_t(\theta) \equiv 1_{\{\theta<t\}},$$

and to abstain otherwise. For each $\gamma \in [0,1]$, we denote by $t_\gamma \equiv F^{-1}(\gamma)$ the γ-quantile of θ, so that the cutoff mechanism π_{t_γ} recommends to consume with probability γ as $\mathbf{E}[\pi_{t_\gamma}(\theta)] = \gamma$.

Lemma 2. *The following holds*:

1. *Among all mechanisms π with $\mathbf{E}[\pi(\theta)] = \gamma$, the cutoff mechanism π_{t_γ} minimizes $\mathbf{E}[\theta\pi(\theta)]$.*
2. *If a mechanism π with $\mathbf{E}[\pi(\theta)] = \gamma$ is IC, then π_{t_γ} is IC as well.*

The intuition of this result is that cutoff mechanisms have good efficiency properties because they recommend to consume for values of θ such that consumption is the most valued by the DM. Moreover, they have good incentive properties because they recommend to abstain when the news about θ is the most alarming.

The two parts of Lemma 2 together imply that designing the optimal IC mechanism boils down to finding the optimal cutoff t. There is a case distinction. The objective in (7) is maximized by the unconstrained-optimal mechanism π_{t^h} that recommends to consume when the net benefit $t^h - \theta$ from consuming is strictly positive. If π_{t^h} is IC, then it solves (7). Otherwise, Constraint (6) becomes binding.

Proposition 1. *If*

$$\mathbf{E}[\theta|\theta \geq t^h] \geq t^a, \tag{8}$$

then the optimal IC mechanism is π_{t^h}. Otherwise, the optimal IC mechanism is π_{t^c}, where $t^c \in (t^h, t^a)$ is uniquely defined by

$$\mathbf{E}[\theta|\theta \geq t^c] = t^a. \tag{9}$$

Combining the two cases, the optimal IC mechanism is $\pi_{t^*}(\theta) \equiv 1_{\{\theta<t^*\}}$, where $t^* \equiv \max\{t^h, t^c\}$ and t^c is the cutoff value that satisfies (9), which, for cutoff mechanisms, is equivalent to (6) being binding.[7] Thus, harmful consumption takes place under the optimal IC mechanism if and only if (8) does not hold. The key insight of (9) is that, following the recommendation to abstain, the DM is on the verge of falling into the harmful-consumption trap (t^h, t^a). This is because the DM's mean posterior belief about θ is at the critical level t^a and is thus just high enough to induce abstention. This achieves an optimal balance between credibility and efficiency: a lower cutoff would undermine the credibility of the mechanism, whereas a higher cutoff would render the recommendation to abstain inefficiently alarming.

Notice that the optimal information nudge only warns the high-risk types. Potentially, a sizable mass of somewhat lower-risk types would prefer a warning as well. Yet the optimal nudge has to sacrifice them in order to convince at least the high-risk types to abstain. An example of such selective nudging are alcohol warnings that target pregnant women instead of the entire population of people who should better drink less.

There are several ways of implementing the optimal IC mechanism: for example, consumption for types $\theta < t^c$ can indifferently be triggered by fully disclosing these types, or by sending the message that $\theta < t^c$. Thus, the optimal information nudge does not have to be simple—but it can be. What is crucial is the composition of the group that receives a warning.

3.2. The Benefits of Optimal Information Design: An Example

To develop an intuition for the potential benefits of optimal information design, suppose that θ is uniformly distributed over $[\underline{\theta}, 1]$ for some $\underline{\theta}$, which, in line with Assumption 1, we allow to vary in $[0, t^h)$. For concreteness, we set the other parameters to $\beta \equiv 1/2$, $\delta \equiv 1$, $C \equiv 3$, so that the welfare-optimal cutoff

for consumption is $t^h = 1/2$, and the behavioral cutoff is $t^a = 2/3$. The optimal information nudge recommends to consume when $\theta < t^h$, and, hence, involves no harmful consumption; indeed, the corresponding recommendation to abstain reveals that $\theta \geq t^h = 1/2$, leading to a mean posterior belief $\hat{\boldsymbol{\theta}} = 3/4 > t^a$, which makes it IC.

Figure 2 compares the date-0 expected utilities from no information, full information, and the optimal information nudge for different values of $\underline{\theta} \in [0, t^h)$. With no information, the DM consumes if $\mathbf{E}[\theta] < t^a$, which holds for $\underline{\theta} < 1/3$. Consumption is most harmful for $\underline{\theta}$ just below that threshold, whereas, for $\underline{\theta} \geq 1/3$, the DM obtains his abstention utility of zero. With full information, the picture is flipped. The gains from consumption outweigh the losses for $\underline{\theta} < 1/3$, whereas, for $\underline{\theta} > 1/3$, harmful consumption dominates and expected utility is negative. The optimal information nudge completely avoids harmful consumption in this example. Moreover, the gains from careful information design are greatest in the critical cases around the harmful-consumption trap. For $\underline{\theta}$ close to zero, the expected utility from the full-information mechanism is not much lower than that from the optimal mechanism because consumption is mostly welfare enhancing. For $\underline{\theta}$ approaching t^h, abstention is guaranteed when receiving no information is an option. The greatest gains from optimal design arise in between these two extremes, for instance, around $\underline{\theta} = 1/3$, where the welfare ordering between no and full information is reversed.

4. Comparative Statics

In this section, we analyze and give intuitions for comparative statics in the risk distribution and the severity of the present bias. The proofs of all the results of this section are provided in the online appendix.

4.1. Changes in the Distribution of Risks

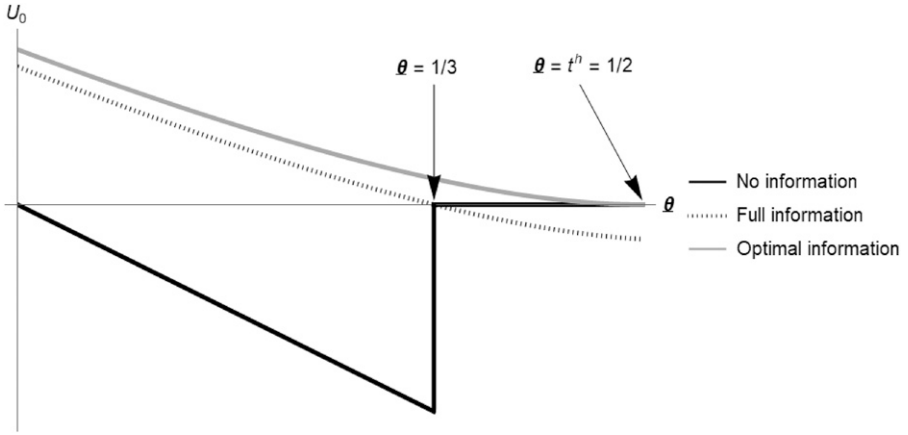
For comparative statics in the distribution of risk, the more interesting scenario arises when (8) does not hold, so that the optimal IC mechanism $\pi_{t^*} = \pi_{t^c}$ involves harmful consumption and $t^* = t^c > t^h$ is the unique solution to (9). In that case, the characterization (9) of the cutoff t^c leads to straightforward comparative statics in terms of the distribution $\mathbf{P}$; for simplicity, we shall assume that all distributions of risks have full support over $[0,1]$. Suppose, for instance, that $\overline{\mathbf{P}}$ dominates $\underline{\mathbf{P}}$ in the hazard-rate order, that is,

$$\frac{\underline{f}(t)}{1-\underline{F}(t)} \geq \frac{\overline{f}(t)}{1-\overline{F}(t)} \text{ for all } t \in [0,1).$$

By the full support assumption, the conditional distributions $\overline{\mathbf{P}}[\cdot|\theta \geq t]$ and $\underline{\mathbf{P}}[\cdot|\theta \geq t]$ are well-defined for all $t \in [0,1)$, and the assumption that $\overline{\mathbf{P}}$ dominates $\underline{\mathbf{P}}$ in the hazard-rate order is equivalent to $\overline{\mathbf{P}}[\cdot|\theta \geq t]$ first-order stochastically dominating $\underline{\mathbf{P}}[\cdot|\theta \geq t]$ for any such t (Shaked and Shanthikumar (2007, section 1.B.1)), which, in turn, implies that $\overline{\mathbf{E}}[\theta|\theta \geq t] \geq \underline{\mathbf{E}}[\theta|\theta \geq t]$. By (9), the cutoff t^c is thus lower under $\overline{\mathbf{P}}$ than under $\underline{\mathbf{P}}$, $\overline{t}^c \leq \underline{t}^c$.

Thus, if the optimal IC mechanism under $\underline{\mathbf{P}}$ involves no consumption for some type, then neither does the optimal IC mechanism under $\overline{\mathbf{P}}$. Intuitively, for any cutoff $t \in [0,1)$, revealing that $\theta \geq t$ is more efficient at discouraging consumption under $\overline{\mathbf{P}}$ than under $\underline{\mathbf{P}}$. Hence, it is credible to choose a lower cutoff t^c under $\overline{\mathbf{P}}$ than under $\underline{\mathbf{P}}$, enabling the mechanism designer to neutralize a larger set of types for which consumption would be harmful. Such types are more likely under $\overline{\mathbf{P}}$ than under $\underline{\mathbf{P}}$ by first-order stochastic dominance. Finally, notice that cases where (8) holds can be discussed in a similar way as t^a does not depend on the distribution of θ and $\overline{t}^c \leq \underline{t}^c$ implies $\overline{t}^* = \max\{t^h, \overline{t}^c\} \leq \max\{t^h, \underline{t}^c\} = \underline{t}^*$. In fact, in line with Kolotilin (2015), we obtain the following stronger result.

Figure 2. Expected Utilities in Different Information Scenarios for Varying Support $[\underline{\theta}, 1]$



Corollary 1. *If* $\overline{\mathbf{P}}$ *dominates* $\underline{\mathbf{P}}$ *in the increasing convex order, that is, if* $\mathbf{E}^{\overline{\mathbf{P}}}[h(\theta)] \geq \mathbf{E}^{\underline{\mathbf{P}}}[h(\theta)]$ *for all nondecreasing and convex* h*, then the probability of consumption is lower under* $\overline{\mathbf{P}}$ *than under* $\underline{\mathbf{P}}$.

Intuitively, two effects are reinforcing each other here: it is more desirable to discourage consumption under $\overline{\mathbf{P}}$ than under $\underline{\mathbf{P}}$, and it is also easier for the mechanism designer to do so, because the optimal abstinence cutoff under $\underline{\mathbf{P}}$ is a fortiori IC under $\overline{\mathbf{P}}$.

4.2. Changes in Present Bias

We now turn to the comparative statics with respect to the severity of the DM's present bias, which is inversely related to β. We start with the basic observation that the cutoff $t^* = \max\{t^h, t^c\}$ for θ above which abstinence is recommended is strictly decreasing in β. Indeed, if (8) holds, then $t^* = t^h \geq t^c$, and this directly follows from (3)–(4); if (8) does not hold, then $t^* = t^c > t^h$, and this directly follows from (3) and (9). Thus, if the optimal IC mechanism for a time-inconsistency parameter $\underline{\beta}$ involves abstinence for a given value of θ, then so does the optimal IC mechanism for any time-inconsistency parameter $\overline{\beta} > \underline{\beta}$. That is, a more severe present bias induces a higher probability of consumption, which corresponds to a tightening of the target group receiving a credible warning to abstain.

We now turn to the more subtle question of how a change in β affects the probability of harmful consumption. We start with a closer examination of Condition (8), under which the unconstrained-optimal mechanism π_{t^h} is IC. By (3)–(4), this condition amounts to

$$\mathbf{E}\left[\theta \,\middle|\, \theta \geq \frac{1+\beta\delta}{(1+\delta)\beta\delta^2 C}\right] \geq \frac{1}{\beta\delta^2 C}. \tag{10}$$

Because a time-consistent DM never engages into harmful consumption, a natural guess is that the optimal IC mechanism involves no harmful consumption and, hence, coincides with π_{t^h} when the DM's present bias is not too severe. This intuition is confirmed by the observation that because the distribution $\mathbf{P}$ has full support, a sufficient condition for (10) to hold is that β be close enough to 1. By assuming some additional regularity on $\mathbf{P}$, we can turn this sufficient condition into an equivalence. Specifically, let us assume that $\mathbf{P}$ satisfies the strict monotone-hazard-rate property (MHRP):

$$\frac{f(t)}{1-F(t)} \text{ is strictly increasing in } t \in [0,1).$$

The following result then holds.

Corollary 2. *If* $\mathbf{P}$ *satisfies MHRP, then* π_{t^h} *is IC if and only if* $\beta \geq \beta^u$*, where* β^u *is the unique value of* $\beta \in (1/(\delta^2 C), 1)$ *that achieves equality in* (10).

Thus harmful consumption takes place if and only if the DM's present bias is severe enough. Now, consider a value of $\underline{\beta} \in (1/(\delta^2 C), \beta^u)$ of β such that harmful consumption takes place under the optimal IC mechanism. Does a small increase in β to some value $\overline{\beta} \in (\underline{\beta}, \beta^u)$ necessarily involve less harmful consumption? There are two opposite effects at play here. On the one hand, from the reasoning at the beginning of this section, there are values of θ such that the DM would be trapped in harmful consumption under $\underline{\beta}$ but abstains under $\overline{\beta}$; on the other hand, according to (3)–(4), the lower bound t^h of the harmful-consumption trap (t^h, t^a) is lower under $\overline{\beta}$ than under $\underline{\beta}$, because the DM attaches greater importance to the harmful consequences of consumption if he has more self-control. The first effect tends to reduce the harmful-consumption trap (t^h, t^c); the second, to increase it. Hence, any statement about how harmful consumption varies with β under the optimal IC mechanism is necessarily of a probabilistic nature. The following result is a first step in that direction. It shows that, under a strengthening of MHRP, harmful consumption is more likely to take place under a more severe present bias.

Corollary 3. *If* $\mathbf{P}$ *satisfies MHRP and its density* f *is such that*

$$\text{for all } t > t', f(t) > \frac{1}{1+\delta} f(t'), \tag{11}$$

then the probability $F(t^c) - F(t^h)$ *that harmful consumption takes place under the optimal IC mechanism is strictly decreasing in* $\beta \in (1/(\delta^2 C), \beta^u)$.

Condition (11) requires that the density f does not decrease too fast over $[0,1]$, so that the margin of risk above which abstinence can be sustained, t^c, remains in a probabilistic sense more important than that above which consumption becomes harmful, t^h. The condition is satisfied, for instance, if $\mathbf{P}$ is the uniform distribution. However, it is not satisfied, for instance, if $\mathbf{P}$ is a Beta(a, b) distribution with $a, b > 1$, which satisfies MHRP, but for which $f(1) = 0$. Corollary A.1 in the online appendix shows that Corollary 3 does not extend to this case. Specifically, whenever the DM's present bias is already severe, a decrease in this bias can actually lead to an increase in the probability of harmful consumption.

5. Traffic-Light Nudges

In practice, not all individuals suffer from the same self-control problems. On the one hand, people seem to differ in their overall self-control capacities (Sutter et al. 2013, Mischel 2014). On the other hand, the specific context matters a lot: whereas smoking may be very tempting for some individuals, others may find

it easy to resist cigarettes, yet lose their self-control when it comes to chocolate. In this section, we analyze which insights from our basic model carry over to the more realistic scenario in which the DM's degree of self-control is not known to the mechanism designer. In particular, we show that in many situations traffic-light nudges with three distinct signal realizations are optimal.

To address these issues, we analyze optimal information nudges in a mixed population, a share $p_L \in (0,1)$ of which has low self-control, and the remaining share p_H has high self-control, with corresponding time-inconsistency parameters $0 < \beta_L < \beta_H \leq 1$. The utility indices of the date-0 and date-1 selves of a DM of type $i = L, H$ are

$$U_{i,0}(x_0, x_1, \theta) \equiv x_0 + \beta_i \delta x_1 - \beta_i \delta^2 \theta C x_0 - \beta_i \delta^3 \theta C x_1, \quad (12)$$

$$U_{i,1}(x_0, x_1, \theta) \equiv x_1 - \beta_i \delta C x_0 - \beta_i \delta^2 \theta C x_1. \quad (13)$$

The riskiness θ is assumed to be known to the mechanism designer and the same for each DM, regardless of the DM's degree of self-control. By contrast, whether a given DM has low or high self-control is not known to the mechanism designer, whose goal is to maximize social welfare at date 0. In the following, we focus on the case where each DM is offered the same information structure, or *joint mechanism*. As a result, both types of DMs are exposed to the same information, as in the case of tobacco, alcohol, and food warnings. For simplicity, we assume that $\Theta = [0,1]$ and that $\mathbf{P}$ satisfies MHRP.

It is clear from (12)–(13) that, for any mean posterior belief $\hat{\boldsymbol{\theta}}$, type L consumes whenever type H does. Therefore, we can focus on direct joint mechanisms $\pi \equiv (\pi_{LH}, \pi_L, \pi_\emptyset)$, where $\pi_j : \Theta \to [0,1]$ associates to each θ the respective probability of issuing a recommendation for both types L and H to consume ($j = LH$), for only type L to consume ($j = L$), or for both types L and H to abstain ($j = \emptyset$). For each $i = L, H$, we denote by t_i^a, t_i^h, t_i^c, and $t_i^* \equiv \max\{t_i^h, t_i^c\}$ the relevant cutoffs defined in Sections 2 and 3 for type i's individually optimal mechanism.

5.1. When the Individually Optimal Mechanisms Do Not Interfere

We first characterize the circumstances in which the individually optimal mechanisms do not interfere, in the sense that they are simultaneously implementable. For every type $i = L, H$, the corresponding mechanism recommends to consume if and only if $\theta < t_i^*$, and we have $t_H^* < t_L^*$. The same outcome can be achieved in a mixed population if and only if the joint mechanism

$$(\pi_{LH}^*, \pi_L^*, \pi_\emptyset^*)(\theta) \equiv \left(1_{\{\theta < t_H^*\}}, 1_{\{t_H^* \leq \theta < t_L^*\}}, 1_{\{\theta \geq t_L^*\}}\right) \quad (14)$$

obtained by merging these mechanisms is IC. By inspection, this is the case if and only if, upon receiving recommendation L, type H is still willing to abstain, that is,

$$\mathbf{E}[\theta \,|\, t_H^* \leq \theta < t_L^*] \geq t_H^a. \quad (15)$$

The following result shows that (15) holds if and only if β_H is sufficiently larger than β_L, so that t_L^* is sufficiently larger than t_H^a. The proof, like those of all the results of this section, is provided in the online appendix.

Proposition 2. *If $\mathbf{P}$ satisfies MHRP, then, for each $\beta_L > 1/(\delta^2 C)$, there exists a threshold $\beta_H^{ni}(\beta_L) \in (\beta_L, 1)$ such that the joint mechanism (14) is IC if and only if $\beta_H \geq \beta_H^{ni}(\beta_L)$. The threshold $\beta_H^{ni}(\beta_L)$ is strictly larger than β^u and is strictly increasing in β_L.*

With its two-cutoff structure, the mechanism (14) has a natural interpretation as a monotone traffic-light nudge, whereby the green-yellow-red labels are used to signal low-intermediate-high riskiness. This makes this nudge especially simple to understand, adding to its potential salience.

5.2. When the Individually Optimal Mechanisms Interfere

We next analyze the case where (15) does not hold, so that the individually optimal mechanisms are not simultaneously implementable. In that case, types L and H exert an externality on each other: at least one of them is bound to suffer from the existence of the other. In line with (5)–(6), the joint mechanism π is IC if and only if

$$\frac{\mathbf{E}[\theta \pi_\emptyset(\theta)]}{\mathbf{E}[\pi_\emptyset(\theta)]} \geq t_L^a, \quad (16)$$

$$\frac{\mathbf{E}[\theta \pi_L(\theta)]}{\mathbf{E}[\pi_L(\theta)]} \leq t_L^a, \quad (17)$$

$$\frac{\mathbf{E}[\theta \pi_L(\theta)]}{\mathbf{E}[\pi_L(\theta)]} \geq t_H^a, \quad (18)$$

$$\frac{\mathbf{E}[\theta \pi_{LH}(\theta)]}{\mathbf{E}[\pi_{LH}(\theta)]} \leq t_H^a. \quad (19)$$

Letting $\Pi_L \equiv \pi_{LH} + \pi_L$ and $\Pi_H \equiv \pi_{LH}$ be the probabilities of consumption for type L and type H, respectively, and given the expression (12) for type i's self-0's utility, the mechanism designer's problem becomes[8]

$$\max\left\{\sum_{i=L,H} p_i \beta_i \{t_i^h \mathbf{E}[\Pi_i(\theta)] - \mathbf{E}[\theta \Pi_i(\theta)]\} : \pi \text{ is IC}\right\}. \quad (20)$$

For simplicity, we first focus on the case where types L and H differ enough in their degrees of self-control that their harmful-consumption traps (t_L^h, t_L^a) and (t_H^h, t_H^a) do not overlap.

Assumption 2. $t_H^a < t_L^h$.

Under Assumption 2, conditional on the same posterior belief $\hat{\boldsymbol{\theta}} \in (t_H^a, t_L^h)$, type L at date 0 favors a higher consumption rate than type H at date 1. By (3)–(4), this is equivalent to

$$\beta_H > \beta_H^{no}(\beta_L) \equiv \frac{(1+\delta)\beta_L}{1+\beta_L\delta},$$

so that β_H is sufficiently larger than β_L. This lower bound for β_H is nevertheless consistent with $\beta_H < \beta_H^{ni}(\beta_L)$, in which case, by Proposition 2, the joint mechanism (14) is not IC. Under Assumption 2, designing an IC joint mechanism is straightforward; for instance, the mechanism designer may offer DM-type L his individually optimal information structure, while sacrificing some risk types of type H. Specifically, in analogy with (9), let $\hat{t}_{LH}(t_L^*) > t_H^*$ be the cutoff uniquely defined by the condition

$$\mathbf{E}[\theta | \hat{t}_{LH}(t_L^*) \le \theta < t_L^*] = t_H^a. \tag{21}$$

Then, the joint mechanism

$$(\pi_{LH}, \pi_L, \pi_\emptyset)(\theta) \equiv \left(1_{\{\theta < \hat{t}_{LH}(t_L^*)\}}, 1_{\{\hat{t}_{LH}(t_L^*) \le \theta < t_L^*\}}, 1_{\{\theta \ge t_L^*\}}\right) \tag{22}$$

is IC because, by construction, (18) is binding. Intuitively, the joint mechanism (22) lets type H abstain as much as possible while providing type L with his individually optimal nudge and maintaining incentive-compatibility for type H. Notice that this mechanism is again a monotone traffic-light nudge. The upshot of this example is that there are gains from pooling intermediate values of θ into a yellow warning. Indeed, the central result of this section more generally states that, under Assumption 2, a two-cutoff joint mechanism is optimal.

Proposition 3. *Under Assumption 2, there exist two cutoffs* $0 < t_{LH}^{**} \le t_L^{**} \le 1$ *such that*

$$(\pi_{LH}^{**}, \pi_L^{**}, \pi_\emptyset^{**})(\theta) \equiv \left(1_{\{\theta < t_{LH}^{**}\}}, 1_{\{t_{LH}^{**} \le \theta < t_L^{**}\}}, 1_{\{\theta \ge t_L^{**}\}}\right) \tag{23}$$

is an optimal IC joint mechanism.

When $t_{LH}^{**} < t_L^{**}$, the optimal IC joint mechanism can be implemented by a three-label monotone traffic-light nudge; as shown in Lemma 3, this inequality always holds under Assumption 2. Low-risk DMs with $\theta < t_{LH}^{**}$ receive a green light to consume regardless of their degree of self-control. Intermediate-risk DMs with $t_{LH}^{**} \le \theta < t_L^{**}$ receive a yellow warning, which is strong enough to induce DMs with high self-control to abstain, though not DMs with low self-control. High-risk DMs with $\theta \ge t_L^{**}$ receive a red warning to abstain regardless of their degree of self-control. Such a monotone traffic-light nudge has an easy-to-grasp connotation.[9] Koenigstorfer et al. (2014) confirm this prediction in an empirical study, comparing consumers with high and low degrees of self-control. Thorndike et al. (2014) and Reisch and Sunstein (2016) document that traffic-light labels are effective in practice.

Proposition 3 generalizes the optimality of cutoff mechanisms to the more realistic case of heterogenous degrees of self-control. In line with Lemma 2 for the homogeneous case, the intuition is based on a comparison of all mechanisms that assign the same probabilities to the same recommendations. As before, using a cutoff t_{LH}^{**} to distinguish between the green light and the yellow warning is good for both efficiency and incentive-compatibility purposes. For the optimal decision whether to display a yellow or a red warning, however, a novel tradeoff arises. On the one hand, pooling the highest-risk types into the red rather than into the yellow warning is good for efficiency purposes because the red warning induces a DM to abstain regardless of his degree of self-control. On the other hand, pooling the highest-risk types into the yellow rather into than the red warning is good for incentive purposes because this relaxes the incentive constraint (18) of type H. In the online appendix, we prove that, under Assumption 2, the first effect dominates, giving rise to a monotone traffic-light nudge. Our next result explicitly characterizes the optimal cutoffs t_{LH}^{**} and t_L^{**} corresponding to the green-yellow and yellow-red boundaries, respectively.

Lemma 3. *Suppose that* (15) *does not hold, so that the individually optimal mechanisms with cutoffs* t_H^* *and* t_L^* *are not simultaneously implementable, and let* $\hat{t}_{LH}(t_L^*)$ *be implicitly defined by* (21)*. Then, the optimal cutoffs* (t_{LH}^{**}, t_L^{**}) *in* (23) *satisfy* $t_{LH}^{**} < t_L^{**}$ *and are given by*

1. $(\hat{t}_{LH}(t_L^*), t_L^*)$ *if and only if*

$$\frac{p_H\beta_H}{p_L\beta_L}\frac{\hat{t}_{LH}(t_L^*) - t_H^h}{t_L^* - t_L^h} \le \frac{t_H^a - \hat{t}_{LH}(t_L^*)}{t_L^* - t_H^a}; \tag{24}$$

2. $(t_H^*, 1)$ *if and only if*

$$\frac{p_H\beta_H}{p_L\beta_L}\frac{t_H^* - t_H^h}{1 - t_L^h} \ge \frac{t_H^a - t_H^*}{1 - t_H^a}; \tag{25}$$

3. *and otherwise the unique solution to*

$$\mathbf{E}[\theta | t_{LH}^{**} \le \theta < t_L^{**}] = t_H^a \quad \text{and} \quad \frac{p_H\beta_H}{p_L\beta_L}\frac{t_{LH}^{**} - t_H^h}{t_L^{**} - t_L^h} = \frac{t_H^a - t_{LH}^{**}}{t_L^{**} - t_H^a}. \tag{26}$$

Unlike cases 1 and 2, which correspond to corner solutions, case 3 corresponds to an interior solution; the standard result then holds that the mechanism designer's marginal rate of substitution equals her marginal cost ratio, where costs are measured in terms of tightening the incentive constraint (18) of type H.

Specifically, the characterization reflects the tradeoff faced by the mechanism designer when she attempts to simultaneously persuade both types of DMs. Pooling marginally more risks into the yellow warning than into the green light by decreasing t^{**}_{LH} comes at a benefit proportional to $p_H\beta_H(t^{**}_{LH} - t^h_H)$ due to higher abstinence of type H. Yet there is also the marginal cost of tightening Constraint (18) from below, which is proportional to $t^a_H - t^{**}_{LH}$. Similarly, pooling marginally more risks into the red rather than into the yellow warning by decreasing t^{**}_L comes at a benefit proportional to $p_L\beta_L(t^{**}_L - t^h_L)$ due to higher abstinence of type L. Yet there is also the marginal cost of tightening Constraint (18) from above, which is proportional to $t^{**}_L - t^a_H$. Balancing these marginal benefits and costs while binding the incentive constraint (18) of type H yields (26).

Finally, the cutoff characterization conditions (24)–(25) enable us to derive straightforward comparative statics with respect to the population share of type H, which determines which of cases 1–3 in Lemma 3 arises.

Corollary 4. *Suppose that* (15) *does not hold, so that the individually optimal mechanisms with cutoffs t^*_H and t^*_L are not simultaneously implementable. Then there exist thresholds $0 \le \underline{p} < \overline{p} \le 1$ such that*

1. *for $p_H \in [0, \underline{p}]$, the optimal IC joint mechanism implements the individually optimal cutoff t^*_L for type L and the cutoff for type H is determined by* (21);
2. *for $p_H \in [\overline{p}, 1]$, the optimal IC joint mechanism implements the individually optimal cutoff t^*_H for type H, whereas type L always consumes*;
3. *for $p_H \in (\underline{p}, \overline{p})$, the optimal IC joint mechanism implements the interior solution to* (26). *Consumption of type H is strictly decreasing in p_H, whereas consumption of type L is strictly increasing in p_H.*

Moreover, $\underline{p} = 0$ if and only if the individually unconstrained-optimal mechanism for type L is IC in the sense of Proposition 1, and similarly $\overline{p} = 1$ if and only if the individually unconstrained-optimal mechanism for type H is IC in the sense of Proposition 1.

We conclude this section with a short discussion of what can happen when Assumption 2 does not hold. In that case, pooling the highest-risk types into the yellow rather than into the red warning may be so efficient at relaxing (18) that it becomes optimal to pool intermediate and extreme values of θ into the yellow warning. A nonmonotone traffic-light nudge may be optimal, but we should stress that such a nudge loses much of the intuitive appeal of those we have encountered so far. Alternatively, if the two types are very similar, a pooling outcome can emerge, in which both types face the individually optimal information nudge for type L.

Proposition 4. *In general, there exist three cutoffs $0 < t^{**}_{LH} \le \underline{t}^{**}_L \le \overline{t}^{**}_L \le 1$ such that*

$$(\pi^{**}_{LH}, \pi^{**}_L, \pi^{**}_\emptyset)(\theta) \equiv \left(1_{\{\theta < t^{**}_{LH}\}}, 1_{\{t^{**}_{LH} \le \theta < \underline{t}^{**}_L\}} + 1_{\{\theta \ge \overline{t}^{**}_L\}}, 1_{\{\underline{t}^{**}_L \le \theta < \overline{t}^{**}_L\}}\right) \tag{27}$$

is an optimal IC joint mechanism.

5.3. A Remark on the Continuous-Type Case

When there is more heterogeneity in the degrees of self-control in the population, incentivizing all types of DMs with a single nudge becomes increasingly difficult. Intuitively, this is because a nudge that is just strong enough to convince a certain β-type to abstain will be just too weak to induce a slightly lower β-type to abstain. In this section, we study how the structure of optimal information nudges changes when the degree of self-control β is continuously distributed, with a strictly positive density over the interval $(1/(\delta^2 C), 1)$.

Because there is a one-to-one correspondence between β and the cutoff $t^a = 1/(\beta\delta^2 C)$, it is convenient for now to think of a DM's private type as being described by $t^a \in [t_0, 1] \equiv [1/(\delta^2 C), 1]$ rather than by β; a high value of t^a corresponds to a low degree of self-control β. We denote by H the cdf of t^a induced by the distribution of β and by h the corresponding density over $[t_0, 1]$, which we assume to be continuously differentiable over $(t_0, 1)$. In this interpretation, the DM thus draws a cutoff t^a as his private type; the mechanism designer then observes θ but not t^a. The final result of this section then follows along the lines of Kolotilin et al. (2017).

Proposition 5. *If the density h of the distribution of cutoffs t^a is log-concave with $h'(t_0) > 0$,*[10] *then there exists a cutoff $\tilde{t} > t_0$ such that an optimal IC joint mechanism prescribes full disclosure of θ for $\theta < \tilde{t}$ and issues a red warning for $\theta \ge \tilde{t}$, which is an IC recommendation to abstain for DMs with $t^a \le \mathbf{E}[\theta \mid \theta \ge \tilde{t}]$. If, in addition, $h(1) = 0$ and $h'(1) < 0$, then we have an interior solution, $\tilde{t} < 1$.*

When information is continuously distributed both on the mechanism designer's and on the DM's sides, we do not expect a simple three-color traffic light to remain optimal—and, indeed, this is not what we find. Instead, there is still a clear red warning, whereas the yellow and green labels are replaced by a more precise, continuous signal. However, it seems plausible, in light of our previous results, that a traffic-light structure, possibly with more than three labels, remains optimal when the distribution of β is discrete. Whether the relevant information nudge has a traffic-light structure, as in Propositions 2 and 3, or a more continuous structure, as in Proposition 5, thus ultimately hinges on the precision of the statistical

information the mechanism designer has about the DM's degree of self-control.

6. More General Objective Functions

So far, we have focused on a benevolent mechanism designer who maximizes the DM's date-0 utility. We now return to the setting of Section 3 but consider more general objective functions for the mechanism designer. For instance, a lobbyist may want to convince as many people as possible to consume. By contrast, a health authority may want to convince as many people as possible to abstain, disregarding the short-term joy from consumption to focus on the long-run effects of harmful consumption.

Motivated by these considerations, we now analyze more flexible objective functions, with the only restriction that the mechanism designer's utility from consumption is continuous and nonincreasing in the DM's risk type. Specifically, we assume that her utility is of the form

$$V(x_0, x_1, \theta) \equiv x_0 v_0(\theta) + x_1 v_1(\theta) \tag{28}$$

for some continuous and nonincreasing period utility functions v_0 and v_1. This class of objective functions includes the polar cases of a lobbyist with $v_0 = v_1 \equiv 1$ and of a health authority with $v_0 = v_1 \equiv -1$, who only care about the probability of consumption irrespective of θ. It also includes the case of a mechanism designer who does not internalize the DM's present bias, and thus maximizes the expected utility of a DM with $\beta = 1$, or of a mechanism designer who maximizes a weighted sum of the expected utilities of the date-0 and date-1 selves or the expected utilities of overlapping generations of DMs currently at different ages.

As the decision problem of the DM is the same as in Section 3, his consumption decision at each date $t = 0, 1$ is again pinned down by his mean posterior belief $\hat{\boldsymbol{\theta}}$. Hence, a mechanism is IC if and only if it satisfies (5)–(6) and the mechanism designer's realized utility is given by 0 if the DM abstains and by

$$v(\theta) \equiv V(1, 1, \theta) = v_0(\theta) + v_1(\theta)$$

if the DM consumes. In the appendix, we show that Lemma 2 is still valid in this more general environment, so that we can again with no loss of generality restrict attention to cutoff mechanisms π_t, $t \in \Theta$. Moreover, the set of IC cutoffs is an interval $\mathcal{I} \equiv [t^c, t^d]$, where

$$t^c \equiv \inf\{t \in \Theta : \mathbf{E}[\theta \mid \theta \geq t] \geq t^a\}, \tag{29}$$

$$t^d \equiv \sup\{t \in \Theta : \mathbf{E}[\theta \mid \theta < t] \leq t^a\}. \tag{30}$$

The definition of t^c in (29) extends the one given in Section 3, defining t^c as the lowest cutoff such that a recommendation to abstain convinces all types above it. Conversely, t^d is the highest cutoff such that a recommendation to consume convinces all types below it. The cutoffs t^c and t^d are thus the extremal values at which the two incentive constraints (5) and (6) are still satisfied. The interval $\mathcal{I}$ is always nonempty as $t^c < t^a < t^d$. Moreover, one of these two constraints is always trivially satisfied, $t^c = \underline{\theta}$ or $t^d = \overline{\theta}$, depending on whether $\mathbf{E}[\theta]$ is above or below t^a. Thus we can find IC mechanisms that maximize (28) by solving the problem

$$\max\{\mathbf{E}[v(\theta)\pi_t(\theta)] : t \in \mathcal{I}\}.$$

Because v is nonincreasing, solving this problem is immediate. Three cases can arise.

Proposition 6. *The following hold*:

i. *If there exists $t \in \mathcal{I}$ such that $v(t) = 0$, then π_t is an optimal IC mechanism.*

ii. *If $v(t) < 0$ for all $t \in \mathcal{I}$, then π_{t^c} is an optimal IC mechanism.*

iii. *If $v(t) > 0$ for all $t \in \mathcal{I}$, then π_{t^d} is an optimal IC mechanism.*

We can interpret Proposition 6 as follows. First, if the mechanism designer's objective and the consumers' incentives are sufficiently aligned, then the mechanism designer's optimal cutoff is IC. Next, if the mechanism designer favors a sufficiently lower consumption rate than the consumers—as in the case of a health authority—then her optimal cutoff may violate IC in (6). Hence, she must tighten the target group that receives a warning to abstain to achieve IC abstention recommendations for the largest possible population of high-risk types. Finally, if the mechanism designer favors a sufficiently higher consumption rate than the consumers—as in the case of a lobbyist—then her optimal cutoff may violate IC in (5). Hence, she must sacrifice some high-risk types to achieve IC consumption recommendations for the largest possible population of low-risk types. In that case, there may be types trapped in harmful consumption who would have abstained without the nudge. This shows how a present-biased consumer can fall prey to an opportunistic information design. For example, nutritionists argue that by issuing warnings for specific high-risk groups only, many unhealthy foods may still feel appropriate for people of lower-risk type (see, e.g., Fuhrman 2011). A warning to a high-risk group can at the same time function as a justification to continue harmful consumption for lower-risk groups.

7. Concluding Remarks

In this paper, we have studied the optimal design of credible information nudges for consumers with present-biased preferences. When all consumers have the same present bias, we have found that the implementation of optimal information structures is easy in the sense that they are of cutoff type: an optimal information nudge should focus on a specific target risk group and present a signal that is credible to this group.

What makes the design of optimal information nudges challenging is that the target group of the nudge has to be adapted to the severity of the present bias. Populations with a severe present bias need a much more drastic signal in order to avoid harmful consumption. From a liberal designer's perspective, this means that fewer consumers can receive a credible signal to abstain. When consumers have different present biases, the traffic-light structure of the optimal nudge addresses this problem by releasing, in addition to the strong, red warning, a specifically milder, yellow warning. Thus heterogeneity in self-control is a rationale for the traffic-light nudges we observe in practice.

A lobbyist aiming at high consumption rates will provide an information nudge of no impact, or, worse, one that tempts people into consumption who would otherwise abstain. If policymakers overlook or underestimate consumers' self-control problems, such a nudge may seem health concerned when in fact the opposite is the case. Policymakers need to figure in the effects of self-control for the design of information nudges that deter harmful consumption.

Acknowledgments

The authors thank Sandro Ambuehl, Kai Barron, Bruno Biais, Catherine Casamatta, Francesc Dilme, Laura Doval, Jannis Engel, Daniel Garrett, Bertrand Gobillard, Paul Heidhues, Alessandro Ispano, Yves Le Yaouanq, George Loewenstein, Stefano Lovo, Collin Raymond, Frank Rosar, Sebastian Schweighofer-Kodritsch, Roland Strausz, Jean Tirole, and Takuro Yamashita for valuable feedback. The authors also thank seminar audiences at École des Hautes Études Commerciales Paris, Toulouse School of Economics, Universität Bonn, and Universität Freiburg, as well as conference participants at the 2018 Durham University Business School Conference on Mechanism and Institution Design, the 2018 European Association for Research in Industrial Economics Annual Conference, the 2018 European Economic Association Annual Congress, the 2018 HeiKaMax Spring Workshop, the 2018 Verein für Socialpolitik Annual Conference, the 2018 Zentrum für Europäische Wirtschaftsforschung Workshop on Market Design, the 2019 Bavarian Micro Day at Universität Ulm, the 2019 Berlin Industrial Organization Day, the 2019 Nordic Conference on Behavioral and Experimental Economics at Kiel Institut für Weltwirtschaft, and the 2019 Paris Workshop on Signaling in Markets, Auctions and Games for many useful discussions. Anke Greif-Winzrieth, Michelle Hörrmann, Nicola Hüholt, and Christine Knopf provided excellent research assistance.

Appendix. Proofs for Sections 3 and 6

To simultaneously prove Proposition 1 and 6, we study a maximization problem that is slightly more general than (7), namely, for some continuous nonincreasing function v,

$$\max\{\mathbf{E}[v(\theta)\pi(\theta)] : \pi \text{ is IC}\}. \tag{A.1}$$

The incentive constraints are (5)–(6) as in Section 3. We first generalize Lemma 2 by showing that we can with no loss of generality restrict attention to cutoff mechanisms π_t, $t \in \Theta$. To prove this claim, notice first that, for each $\gamma \in [0,1]$, among all mechanisms π with recommendation probability $\mathbf{E}[\pi(\theta)] = \gamma$, the cutoff mechanism π_{t_γ} with cutoff $t_\gamma \equiv F^{-1}(\gamma)$ concentrates as much mass as possible on small values of θ; hence, as v is nonincreasing, it maximizes $\mathbf{E}[v(\theta)\pi(\theta)]$ in this class of mechanisms. Moreover, inspecting the IC conditions (5)–(6), we find that they depend on the mechanism π only through $\mathbf{E}[\pi(\theta)]$ and $\mathbf{E}[\theta\pi(\theta)]$. In particular, holding $\gamma = \mathbf{E}[\pi(\theta)]$ fixed, both constraints only become easier to satisfy when $\mathbf{E}[\theta\pi(\theta)]$ is made smaller, which is again achieved by the cutoff mechanism π_{t_γ}. Thus, if a mechanism π with $\mathbf{E}[\pi(\theta)] = \gamma$ is IC, then π_{t_γ} is IC as well. The claim follows.

We next characterize under which conditions a cutoff mechanism is IC. To this end, notice that for π_t the incentive constraints (5)–(6) can be written as

$$m(t) \equiv \mathbf{E}[\theta \mid \theta < t] \le t^a, \tag{A.2}$$

$$M(t) \equiv \mathbf{E}[\theta \mid \theta \ge t] \ge t^a. \tag{A.3}$$

Under our assumptions on $\mathbf{P}$, both m and M are continuous, strictly increasing functions of $t \in \Theta$, which satisfy $m(\underline{\theta}) = \underline{\theta}$, $m(\overline{\theta}) = \mathbf{E}[\theta]$, $M(\underline{\theta}) = \mathbf{E}[\theta]$, and $M(\overline{\theta}) = \overline{\theta}$. We distinguish two cases. If $t^a > \mathbf{E}[\theta]$, then (A.2) is automatically satisfied, and (A.3) is satisfied if and only if $t \ge t^c$, where $t^c \equiv M^{-1}(t^a) \in (\underline{\theta}, \overline{\theta})$. Alternatively, if $t^a \le \mathbf{E}[\theta]$, then (A.3) is automatically satisfied, and (A.2) is satisfied if and only if $t \le t^d$, where $t^d \equiv m^{-1}(t^a) \in (\underline{\theta}, \overline{\theta}]$. In either case, we denote by $\mathcal{I} \equiv [t^c, t^d]$ the set of IC cutoffs; thus $t^d = \overline{\theta}$ if $t^a > \mathbf{E}[\theta]$ and $t^c = \underline{\theta}$ if $t^a \le \mathbf{E}[\theta]$. Using the Definitions (A.2)–(A.3) of the functions m and M and the definitions of t^c and t^d, it is easy to check that $t^a \in (t^c, t^d)$, as expected. Thus (A.1) boils down to maximizing

$$\mathbf{E}[v(\theta)\pi_t(\theta)] = \int_{\underline{\theta}}^{t} v(\theta) f(\theta)\, \mathrm{d}\theta$$

with respect to $t \in \mathcal{I}$. Solving this problem is immediate given that v is nonincreasing, which implies that the objective function is quasiconcave. If $v > 0$ over $\mathcal{I}$, then $t = t^d$ is the unique solution. If $v < 0$ over $\mathcal{I}$, then $t = t^c$ is the unique solution. Finally, if v vanishes over $\mathcal{I}$, then any $t \in \mathcal{I} \cap v^{-1}(0)$ is a solution. This concludes the proof of Proposition 6.

It remains to complete the proof of Proposition 1. In that case, $v(\theta) \equiv t^h - \theta$ is strictly decreasing and linear in θ, with $v(t^h) = 0$ for $t^h < t^a$. Because the set $\mathcal{I} = [t^c, t^d]$ of IC cutoffs is such that $t^c < t^a < t^d$, we can conclude that $t^h < t^d$, so that the cutoff t^h can never be too high for being IC. Accordingly, when $t^h \ge t^c$, which is exactly (8), we have $t^h \in \mathcal{I}$ and the optimal IC mechanism is π_{t^h}. By contrast, when $t^h < t^c$, which is exactly the negation of (8), we have $t^c > \underline{\theta}$ and the optimal IC mechanism is π_{t^c}, because in that case $v < 0$ over $\mathcal{I}$; finally, (9) is exactly $M(t^c) = t^a$. This concludes the proof of Proposition 1. An important observation is that Constraint (5) is slack at the optimum.

Endnotes

[1] See, e.g., Hammond et al. (2005). Similar findings have been reported regarding alcohol warning labels (MacKinnon et al. 1993) and mandatory calorie posting in chain restaurants (Bollinger et al. 2011).

[2] Harmful consumption is a behavioral property of our model that could be identified in the data by asking subjects whether they would rather always consume or always abstain. The unhappy smoker is a case in point.

[3] Though evidence on the latter is mixed, see VanEpps et al. (2016).

[4] The assumption that $\mathbf{P}$ admits a density is only made to simplify the exposition. When F is discontinuous, the optimal persuasion mechanism may involve mixing at an atom of $\mathbf{P}$.

[5] See Cadario and Chandon (2019) for a literature overview on eating nudges.

[6] See, for example, Caplin and Leahy (2004), Köszegi (2003), and Schweizer and Szech (2018).

[7] By convention, t^c is set equal to $\underline{\theta}$ if there exists no solution to (9), that is, if $\mathbf{E}[\theta] > t^a$.

[8] The population shares p_L and p_H in (20) can also be interpreted as Pareto weights in the social welfare function. This is because we study a pure information-design problem, with no aggregate resource constraint.

[9] A red warning may be especially salient. The empirical literature is mixed on whether traffic-light labels render the provision of information more effective or not, see VanEpps et al. (2016) for a discussion. Yet, of course, this aspect is beyond the analysis of this paper.

[10] If we think of the discount factor β as being generated by a discount rate R, that is, $\beta \equiv 1/(1+R)$, then the log-concavity of h is equivalent to the log-concavity of the density of R. Together with the assumption that $h'(t_0) > 0$ and $h'(1) < 0$, Proposition 5 thus requires that the distribution of R be well-behaved and unimodal.

References

Ainslie G (1975) Specious reward: A behavioral theory of impulsiveness and impulse control. *Psych. Bull.* 82(4):463–496.

Ainslie G (1992) *Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person* (Cambridge University Press, Cambridge, UK).

Alizamir S, de Véricourt F, Wang S (2020) Warning against recurring risks: An information design approach. *Management Sci.* 66(10):4612–4629.

Aprahamian H, Bish DR, Bish EK (2019) Optimal risk-based group testing. *Management Sci.* 65(9):4365–4384.

Argo JJ, Main KJ (2004) Meta-analyses of the effectiveness of warning labels. *J. Public Policy Marketing* 23(2):193–208.

Bénabou R, Tirole J (2002) Self-confidence and personal motivation. *Quart. J. Econom.* 117(3):871–915.

Benkert J-M, Netzer N (2018) Informational requirements of nudging. *J. Political Econom.* 126(6):2323–2355.

Blackwell D (1953) Equivalent comparisons of experiments. *Ann. Math. Statist.* 24(2):265–272.

Bollinger B, Leslie P, Sorensen A (2011) Calorie posting in chain restaurants. Amer. Econom. J. *Econom. Policy* 3(1):91–128.

Brocas I, Carrillo JD (2007) Influence through ignorance. *RAND J. Econom.* 38(4):931–947.

Cadario R, Chandon P (2019) Which healthy eating nudges work best? A meta-analysis of field experiments. *Marketing Sci.* 39(3):465–486.

Caplin A, Leahy J (2004) The supply of information by a concerned expert. *Econom. J.* 114(497):487–505.

Carrillo JD, Mariotti T (2000) Strategic ignorance as a self-disciplining device. *Rev. Econom. Stud.* 67(3):529–544.

Drakopoulos K, Jain S, Randhawa R (2021) Persuading customers to buy early: The value of personalized information provisioning. *Management Sci.* 67(2):828–853.

Fuhrman J (2011) *Eat to Live* (Little Brown, New York).

Ganguly A, Tasoff J (2017) Fantasy and dread: The demand for information and the consumption utility of the future. *Management Sci.* 63(12):4037–4060.

Gao SY, He Y, Zhang R, Zheng Z (2021) Optimizing initial screening for colorectal cancer detection with adherence behavior. Mimeo, Singapore Management University. Accessed May 16, 2022, https://ssrn.com/abstract=3951864.

Golman R, Loewenstein G, Molnar A, Saccardo S (2021) The demand for, and avoidance of, information. *Management Sci.* ePub ahead of print December 15, https://doi.org/10.1287/mnsc.2021.4244.

Gul F, Pesendorfer W (2001) Temptation and self-control. *Econometrica* 69(6):1403–1435.

Gutjahr E, Gmel G, Rehm J (2001) Relation between average alcohol consumption and disease: An overview. *Eur. Addiction Res.* 7(3):117–127.

Habibi A (2020) Motivation and information design. *J. Econom. Behav. Organ.* 169:1–18.

Hall W (2010) What are the policy lessons of national alcohol prohibition in the United States, 1920–1933? *Addiction* 105(7):1164–1173.

Hammond D, Fong GT, McNeill A, Borland R, Cummings KM (2005) Effectiveness of cigarette warning labels in informing smokers about the risks of smoking: Findings from the International Tobacco Control (ITC) four country survey. *Tobacco Control* 15(Suppl III):iii19–iii25.

Hankin JR, Firestone IJ, Sloan JJ, Ager JW, Goodman AC, Sokol RJ, Martier SS (1993) The impact of the alcohol warning label on drinking during pregnancy. *J. Public Policy Marketing* 12(1):10–18.

Ho EH, Hagmann D, Loewenstein G (2020) Measuring information preferences. *Management Sci.* 67(1):126–145.

Kamenica E, Gentzkow M (2011) Bayesian persuasion. *Amer. Econom. Rev.* 101(6):2590–2615.

Kaskutas LA (1993) Changes in public attitudes toward alcohol control policies since the warning label mandate of 1988. *J. Public Policy Marketing* 12(1):30–37.

Koenigstorfer J, Groeppel-Klein A, Kamm F (2014) Healthful food decision making in response to traffic light color-coded nutrition labeling. *J. Public Policy Marketing* 33(1):65–77.

Kolotilin A (2015) Experimental design to persuade. *Games Econom. Behav.* 90:215–226.

Kolotilin A, Mylovanov T, Zapechelnyuk A, Li M (2017) Persuasion of a privately informed receiver. *Econometrica* 85(6):1949–1964.

Köszegi B (2003) Health anxiety and patient behavior. *J. Health Econom.* 22(6):1073–1084.

Laibson D (1997) Golden eggs and hyperbolic discounting. *Quart. J. Econom.* 112(2):443–477.

Lingenbrink D, Iyer K (2019) Optimal signaling mechanisms in unobservable queues. *Oper. Res.* 67(5):1397–1416.

Lipnowski E, Mathevet L (2018) Disclosure to a psychological audience. *Amer. Econom. J. Microeconom.* 10(4):67–93.

Loewenstein G, Prelec D (1992) Anomalies in intertemporal choice: Evidence and an interpretation. *Quart. J. Econom.* 107(2):573–597.

MacKinnon DP, Pentz MA, Stacy AW (1993) The alcohol warning label and adolescents: The first year. *Amer. J. Public Health* 83(4):585–587.

Mazis MB, Morris LA, Swasy JL (1991) An evaluation of the alcohol warning label: Initial survey results. *J. Public Policy Marketing* 10(1):229–241.

McCool J, Webb L, Cameron LD, Hoek J (2012) Graphic warning labels on plain cigarette packs: Will they make a difference to adolescents? *Soc. Sci. Medicine* 74(8):1269–1273.

Miron JA, Zwiebel J (1991) Alcohol consumption during Prohibition. *Amer. Econom. Rev.* 81(2):242–247.

Miron JA, Zwiebel J (1995) The economic case against drug prohibition. *J. Econom. Perspect.* 9(4):175–192.

Mischel W (2014) The Marshmallow Test: Understanding Self-Control and *H*ow to Master It (Little Brown, New York).

Papanastasiou Y, Bimpikis K, Savva N (2018) Crowdsourcing exploration. *Management Sci.* 64(4):1727–1746.

Peleg B, Yaari ME (1973) On the existence of a consistent course of action when tastes are changing. *Rev. Econom. Stud.* 40(3): 391–401.

Phelps ES, Pollak RA (1968) On second-best national saving and game-equilibrium growth. *Rev. Econom. Stud.* 35(2):185–199.

Rayo L, Segal IR (2010) Optimal information disclosure. *J. Political Econom.* 118(5):949–987.

Reisch LA, Sunstein CR (2016) Do Europeans like nudges? *Judgment Decision Making* 11(4):310–325.

Schweizer N, Szech N (2018) Optimal revelation of life-changing information. *Management Sci.* 64(11):5250–5262.

Shaked M, Shanthikumar JG (2007) *Stochastic Orders* (Springer, New York).

Stewart DW, Martin IM (1994) Intended and unintended consequences of warning messages: A review and synthesis of empirical research. *J. Public Policy Marketing* 13(1):1–19.

Strotz RH (1956) Myopia and inconsistency in dynamic utility maximization. *Rev. Econom. Stud.* 23(3):165–180.

Sutter M, Kocher MG, Glätzle-Rützler D, Trautmann ST (2013) Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *Amer. Econom. Rev.* 103(1):510–531.

Szydlowski M (2021) Optimal financing and disclosure. *Management Sci.* 67(1):436–454.

Thaler R (1981) Some empirical evidence on dynamic inconsistency. *Econom. Lett.* 8(3):201–207.

Thaler R, Sunstein CR (2008) *Nudge: Improving Decisions about Health, Wealth, and Happiness* (Yale University Press, New Haven, CT).

Thorndike AN, Riis J, Sonnenberg LM, Levy DE (2014) Traffic-light labels and choice architecture: Promoting healthy food choices. *Amer. J. Preventive Medicine* 46(2):143–149.

VanEpps EM, Downs JS, Loewenstein G (2016) Calorie label formats: Using numeric and traffic light calorie labels to reduce lunch calories. *J. Public Policy Marketing* 35(1):26–36.

Yoon H (2020) Impatience and time inconsistency in discounting models. *Management Sci.* 66(12):5850–5860.

Zimmermann F (2015) Clumped or piecewise? Evidence on preferences for information. *Management Sci.* 61(4):740–753.