International Journal of Research in Marketing 37 (2020) 521–543

Contents lists available at ScienceDirect

IJRM
International Journal of Research in Marketing

journal homepage: www.elsevier.com/locate/ijresmar

Full Length Article

Flying with a net, and without: Preventative devices and self-control☆

Raghunath Singh Rao [a,*], Julie Irwin [a], Zhuping Liu [b]

[a] McCombs School of Business, The University of Texas at Austin, United States of America
[b] Baruch College, City University of New York (CUNY), United States of America

ARTICLE INFO

Article history:
First received on December 9, 2018 and was under review for 6 months
Available online 4 February 2020

Area Editor: Sanjay Jain

Keywords:
Self-control
Present bias
Vice goods
Hyperbolic discounting
Behavioral economics

ABSTRACT

Excessive consumption of many vice goods (e.g., alcohol) has both possible immediate (e.g., a drunk-driving crash) and delayed (e.g., liver disease) negative consequences. This research models the consumption choices of a consumer population with heterogeneous impatience and varying degrees of sophistication in conjunction with both immediate and delayed consequences of excessive vice good consumption. We show that even when a preventative device (e.g., a designated driver) that could eliminate the immediate dangers is available for (almost) free, some consumers forgo it in order to try to use the immediate danger as a soft tool to regulate excessive consumption (e.g., "*If I know I don't have a designated driver, then I won't drink too much.*"). Surprisingly, this "flying without a net" is a successful strategy for some consumers, and we quantify when it is likely to be successful versus harmful. This counterintuitive result has many consequences; e.g., public policies that make the provision of preventative devices compulsory could increase consumer welfare under certain conditions but overall are not Pareto-improving. Likewise, advertising campaigns that exaggerate the likelihood of immediate dangers may be welfare decreasing despite their good intentions. In exploring the effect of our model on pricing of preventative device, we find that pricing is not monotonic in the probability of danger. We also show that consumer pessimism about self-regulation can induce consumers to experience a "boomerang effect" of overconsumption in the presence of a preventative device. Overall our research shows that the relationship between prevention and risk is not as simple as might be assumed in standard analyses and that marketers and public policymakers often need to consider the short-term and the long-term risks as well how prone the consumers are to temptation and how sophisticated they are in judging their temptation.

> In the soul of a person, there is a better part and a worse part. When the naturally better part is in control of the worse, this is what is meant by 'master of himself.'
>
> [(Plato, *Republic*)]

> I have been both master and disciple of myself. I have been a better master than a disciple.
>
> [(Antonio Porchia, Voces)]

☆ We are grateful to the seminar participants at Texas A&M University, Tulane University, Georgetown University and INFORM Marketing Science Conference 2018 for valuable suggestions and feedback.
* Corresponding author.
E-mail addresses: Raghunath.Rao@McCombs.utexas.edu, (R.S. Rao), Julie.Irwin@McCombs.utexas.edu, (J. Irwin), Zhuping.Liu@baruch.cuny.edu. (Z. Liu).

https://doi.org/10.1016/j.ijresmar.2020.01.002
0167-8116/

1. Introduction

Individuals' short-term desires and long-term goals are often in conflict, leading to time-inconsistent behavior. Consumers often plan to eat healthy food but cannot resist junk food sitting in front of a TV; they sign up for expensive gym memberships only to ignore the alarm bell in the mornings; they vow to drink wisely after a severe hangover on a Monday, only to repeat their binge drinking the following weekend. Many of modern society's problems — addictions, crippling debt, unwanted pregnancies, sexually transmitted diseases, the obesity epidemic, domestic violence — have their origins, at least partly, in individuals' inability to self-regulate (Baumeister & Vohs, 2003; Thaler & Shefrin, 1981). Thus, it comes as no surprise that the effects of self-control have attracted increasing attention from academic researchers in domains as diverse as philosophy (Elster, 1979), medicine (Moffitt et al., 2011), and economics (Gul & Pesendorfer, 2004).

Consumers make attempts at self-regulation through a variety of personal means, such as goals and promises made to themselves or others or via self-signaling through virtuous behavior (Carrillo & Dewatripont, 2008; Jain, 2009; Schaefer, Rao, & Mahajan, 2018). Some of this help also takes the form of market goods, such as smaller but expensive packages of unhealthy snacks that help consumers regulate their caloric intake (Jain, 2012; Wertenbroch, 1998) and low-return but illiquid assets that help people to save more (Laibson, 1997). However, for these commitments to work as intended, consumers need to have a good forecast of their future behavior, or else they might end up either forgoing the commitment made or expend resources on more commitment than needed (Heidhues & Kőszegi, 2009; Brocas, Carrillo, & Dewatripont, 2004a).

Self-control issues typically arise when there is a delay between pleasure (i.e., benefits) and consequences (i.e., costs), or vice-versa. Because their immediate desires and long-term goals are often in conflict, people commonly over-consume leisure goods that provide immediate benefits and delayed costs. In the present study, we consider a special category of leisure goods often referred to as "vice goods," wherein excessive consumption not only carries delayed costs but also often comes with potentially *large immediate costs* along with immediate gratification (Loewenstein, 1996). For example, for many people, alcohol is pleasurable in the short term and thus provides an immediate benefit, but excessive and regular consumption of alcohol carries health risks to the liver, brain, and heart over a longer horizon with potentially hazardous consequences (a delayed cost). Additionally, a night of heavy drinking in a bar could also result in a catastrophic road accident (an immediate cost). Many other behaviors fall into this "immediate gratification/immediate cost and delayed costs" category, such as narcotic drugs, fast driving, and casual sexual encounters. To the best of our knowledge, previous work has not simultaneously modeled both the immediate and the delayed consequences of the consumption of vice goods.

Furthermore, much of the existing literature in the self-control domain (e.g., O'Donoghue & Rabin, 1999) has considered consumers who are either unable to forecast overconsumption ("naïve") or are knowledgeable about their tendency to overindulge ("sophisticated"). Naive consumers mistakenly believe that they have no self-control problems (Akerlof, 1991; Phelps & Pollak, 1968) whereas sophisticated consumers can accurately forecast that their future immediate gratification desires would wreak havoc with their best laid out long-term plans (Jain, 2012; Laibson, 1997). Note that naïve consumers perceive commitments to be useless because they believe they do not need them while sophisticated consumers will buy the most appropriate commitment devices, if available, to prevent overconsumption (DellaVigna & Malmendier, 2004). Although these extreme cases provide useful benchmarks, it is reasonable to assume that most people fall somewhere in the middle of the two extremes. Thus, in our model, we consider both consumers' heterogeneity in impatience as well as their awareness about it.

When a (relatively sophisticated) consumer is planning for the future consumption of a vice good, she might want to moderate her consumption to avoid both immediate and delayed costs. At the same time, during planning, she might forecast that when the moment of consumption arrives, she will be tempted to consume more. We include the possibility of the availability of a preventative device (henceforth referred to as a PD) that could take the immediate cost of excessive consumption off the table. For example, a condom acts as a PD as it can eliminate the immediate cost of contracting a sexually transmitted disease. Similarly, a designated driver could eliminate the danger of a drunk driving crash. From a normative perspective, if a PD is available at no cost, then its provision for an agent is almost always welfare enhancing.

On the other hand, when consumers suffer from present bias, from a planning perspective, the immediate cost of excessive immediate gratification might act as a natural barrier that could deter them from excessive consumption, much as a tightrope walker might be extra careful if there were no net below to catch him. To provide another example, an individual planning to engage in moderate consumption of alcohol at an upcoming party might refuse to use an external PD such as a designated driver that would eliminate the danger of drunk driving. He is forecasting that he might be tempted to drink heavily at the party, but if he does not have a designated driver with him, then the fear that drunk driving might result in a catastrophic accident could deter him from excessive drinking. Similarly, someone might decide not to carry a condom to a party if she believes that the fear of contracting STDs or getting pregnant is strong enough to deter her from engaging in casual sex.[1]

[1] The preventative devices we focus upon are different from the commitment devices that have received some attention in the self-control literature (e.g., Carrillo and Dewatripont 2004). We refer to a device as preventative if it prevents the immediate dangers of excessive consumption of vice goods (e.g., a designated driver) even if the excessive consumption does occur. Commitment devices instead help in moderating the consumption of vice goods (e.g., small packages of unhealthy snacks to prevent overeating). Put metaphorically, PDs are like nets under the tightrope and commitment devices are mechanisms to prevent getting on the tightrope in the first place. This distinction is important and we discuss this in the discussion section.

1.1. Related papers

We confine our discussion to a couple of studies that are closest in spirit to ours. In a recent work, Kaur, Kremer, and Mullainathan (2015) asked workers to self-select themselves into one of the two contracts (among other treatments): a "control" contract that paid an incentive wage rate of "b" for every unit of production and a "treatment" contract that paid a wage rate of "b/2" till a target production level and a wage rate of "b" per unit after hitting the target. In a rational world, the latter contract is dominated by the former and hence should never be chosen. But a significant number of workers picked the dominated contracts resulting in an improvement in the productivity among many of these workers. In the language of our paper, the workers picking up the dominated contract willingly put themselves in a "danger" (low wage) to self-regulate and avoid shirking. Jain and Li (2018) is the work from marketing that is closest in spirit to our paper. They study pricing and quality decisions for a firm selling a vice good and find that sometimes a firm can profit from lower quality (higher unhealthiness) and higher prices that allow some consumers to exercise self-control. This idea shares similarity with one of our findings that consumers might willingly put themselves in danger to self-regulate, which is akin to Jain and Li's increased attractiveness of a poorer quality product. However, there are many significant differences between our setup and theirs. Our focus in this paper is not just on self-control problems but also on our proposal that people's lack of perfect awareness of self-control could result in additional problems concerning PDs. For example, in our context picking an inferior product (metaphorically, forgoing the PD) could also fail and decrease consumer welfare. In addition to modeling differences, we focus primarily upon public policy implications, such as the use of Public Service Announcements (PSAs) and compulsory provision of PDs.[2]

1.2. Contributions of this study

Policymakers seem to understand that people often fail to self-regulate their consumption of vice goods and thus, it is understandably popular to advocate the use of PDs to save people from the potential devastating immediate consequences of their lack of control. Our analysis uncovers that the efficacy of consumers' use of PDs is more nuanced and is affected by consumers' self-control as well as their sophistication about this self-control. Our novel model incorporates both the immediate and the delayed consequences of vice goods and captures the heterogeneity in impatience *as well as* consumers' sophistication about self-control. Without this new framework, results such as consumers' willful forgoing of a free PD seems patently non-normative. Furthermore, our model allows us to analyze the effects of policies relevant to vice goods and PDs, including those highlighting immediate dangers (e.g., PSAs about the dangers posed by the consumption of vice goods), and the policies that exaggerate the probability of such dangers (e.g., advertising showing deaths from a drug overdose). We show that in contrast with conventional wisdom, these policies always come at the cost of the reduction in the welfare of a subset of consumers.[3] We also study the pricing policy of a monopolist seller of a PD and show that its price is non-monotonic in the probability of danger. Finally, our model analysis can speak to *intrapersonal* strategies like strategic pessimism that seek to overcome the self-regulation problems arising from overconfidence.

We organize the rest of the article as follows: In Section 2, we describe the baseline model followed by main analysis and results in Section 3 and present three extensions of our model in Section 4. We conclude with the discussion and implications of our findings in Section 5.

2. Base model

Consider a consumer population (who are considering a vice good consumption) who discount time by $0 \le \delta \le 1$. The consumers face self-control issues and, as is standard, we denote their short-run discounting parameter (impatience) by β with $0 \le \beta \le 1$. We employ the widely used quasi-hyperbolic discounting approach (Thaler & Shefrin, 1981) to model consumers, which has been used in a wide variety of contexts such as consumption-saving decisions (Ashraf, Karlan, & Yin, 2006), DVD viewing patterns (Milkman, Rogers, & Bazerman, 2009) and mail-in rebates (Gilpatric, 2009). Consistent with the overconfidence literature, we assume that consumers are often unable to forecast their present bias: the parameter $\hat{\beta}$ represents consumers' belief about their impatience with $\beta \le \hat{\beta} \le 1$. Hence, $(\hat{\beta} - \beta)$ represents consumers' *overconfidence* in their self-regulation, and each consumer is represented by a unique coordinate $(\hat{\beta}, \beta)$. At any time t, the NPV for a stream of utilities (with per-period utility denoted by u_t) is given by (DellaVigna, 2009; Laibson, 1997):

$$U_t = u_t + \beta\delta u_{t+1} + \beta\delta^2 u_{t+2} + \beta\delta^3 u_{t+3} + \dots \tag{1a}$$

Here, $\beta < 1$ represents the self-control parameter. The discounting between any two subsequent periods in the future (from t's perspective) is denoted by δ and the discounting between the present and the next period is sharper at $\beta\delta$. The lower the value of β, the more severe is the "present bias." Although Eq. (1a) determines the consumer's choices if he were to pick them at time t, at each period after t, he will again suffer from the same present-bias that plagues him at time t. Consumers might have some

[2] Other technical differences with Jain and Li (2018) are highlighted at appropriate places within the text.

[3] In a simple extension, we also show that allowing for price of the vice good to be endogenously determined does not lead to market overcoming the efficiencies induced by people's self-regulation problems- in fact, they might be exacerbated in some instances.

Table 1
Model notations.

Symbol	Description
δ	Long-run (exponential) discount factor
β	Short-run discount factor
$\hat{\beta}$	Perceived short-run discount facto
U_H	Utility from excessive consumption
U_L	Utility from moderate consumption
ΔU	Degree of temptation ($U_H - U_L$)
p	Probability of short-term danger
$\hat{p}$	Perceived probability of short-term danger
I	Cost of short-run danger
T	Cost of obtaining preventative device
C_H	Delayed long-term cost of vice consumption
$\hat{C}_H$	Perceived long-term cost
e	Externality imposed due to excessive consumption (without a PD)
τ	PD adoption rate under enforcement

awareness about this "future" present-bias, which for any $s \geq 1$ at time t is modeled as (O'Donoghue & Rabin, 1999):

$$\hat{U}_{t+s} = u_{t+s} + \hat{\beta}\delta u_{t+s+1} + \hat{\beta}\delta^2 u_{t+s+2} + \hat{\beta}\delta^3 u_{t+s+3} + \ldots \tag{1b}$$

As we noted before, $\hat{\beta} \geq \beta$. A consumer could be *sophisticated* about her future self-control ($\hat{\beta} = \beta$), *perfectly naïve* ($\hat{\beta} = 1$), or *partially naïve* ($1 > \hat{\beta} > \beta$) about her self-regulation.

To focus upon the self-control issues (e.g., as in Carrillo & Dewatripont, 2008; Jain & Li, 2018), we assume δ to be the same across all consumers and normalize it to 1. To specialize the model to our context, we embed several features. To capture heterogeneity in consumers' actual short-run discounting (β) and their predicted short-term discounting ($\hat{\beta}$), we represent the population by the upper triangle in Fig. 2. The consumers along the diagonal have $\beta = \hat{\beta}$, and are *perfectly sophisticated* in their forecast of impatience, whereas the consumers on the upper horizontal line with ($\beta < 1, \hat{\beta} = 1$) are *perfectly naïve* and mistakenly believe that they have no self-control problems. The rest of the consumers inside the triangle are *partially naïve* (or *partially sophisticated*) —that is, imperfectly aware of their time inconsistency. Furthermore, we add immediate costs that accompany the consumption of vice goods in addition to the delayed cost in the standard model. Table 1 provides a summary of the notations used in the model.

2.1. Benchmark: no preventative device (PD) available

We consider a three-period model in which the period 0 consists of planning when the consumer is anticipating her consumption, period 1 consists of consumption, and period 2 consist of the realization of the delayed costs. As a benchmark, initially, we assume no PD is available. Consumers' decision process is illustrated in Fig. 1.

A consumer gets an instantaneous utility of U_H from excessive consumption and U_L from moderate consumption[4] with, $U_H > U_L$: that is, absent any costs, high consumption is always preferred to moderate consumption. Denote $U_H - U_L \equiv \Delta U$ which represents the strength of temptation facing consumers. A person who consumes high could potentially suffer from an immediate danger I with probability p. Also, she suffers from a delayed cost of C_H. We normalize both the short-term and long-term costs of moderate consumption to be zero.[5] Furthermore, we assume that $\Delta U - pI > 0$, which means that just the short-term danger in itself will not deter a consumer from overconsumption (but long-term harm could- see below)

In period 0, the utility of high and moderate consumption in period 1 is given respectively by (using Eq. (1a)):

$$U_0(H) = \beta(U_H - pI) - \beta C_H \text{ and}, U_0(L) = \beta U_L \tag{2}$$

Before proceeding further, it is useful to consider the formulation of excessive consumption and the associated short-term harm in the model. Since the high consumption of the vice good results in a utility of $U_H - pI$, and furthermore the short-term harm is additively separable; one could potentially argue that "*pI*" could just be part of the quality formulation. However, there are important subtle differences between that modeling choice and the one we made here. First, short-term harm arises as *a consequence* of the high consumption of the vice-good, and as such is distinct from the quality of the product. As an example, the quality of an alcoholic drink is its taste and the associated enjoyment while the probability of a blackout or getting into an

[4] The utilities are assumed to be net of the prices paid for the respective consumption choices. We relax this later.

[5] In general, there is a temporal distance between consumption and short-term harm and further temporal distance between (potential) short-term harm and long-term harm. While in principle one could create a model that temporally separates consumption and (potential) short-term harm, this addition would create substantial complexity to the model (with four-periods) that, for this project, we do not feel is worth the added explication given our overall set-up. Our basic assumption is that short-term injury is very close to the consumption (i.e., the accident or drug overdose could happen the evening of drinking) while the long-term damage is potentially much later in the future (i.e., liver cancer or mental health issues would appear months or years later). We thank AE for bringing this to our attention.

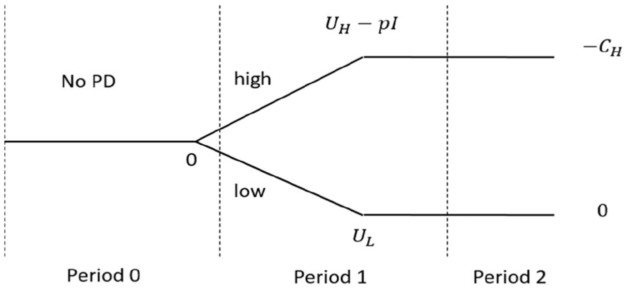

Fig. 1. Timing with no preventative device.

accident is short-term harm. The former is product specific while the latter is more situational. Second, an individual can potentially avoid short-term harm if they use a preventative device – in that sense, this feature makes it distinct from the quality. This distinction suggests a role for policy- the main focus of our paper. In other words, in our set-up a policymaker can design interventions that could potentially affect only the short-term dangers without affecting the associated quality.[6]

To focus on the interesting cases wherein the consumers' different (i.e., time-inconsistent) selves might conflict, we assume that:

$$\Delta U < C_H \rightarrow U_0(H) < U_0(L). \tag{3}$$

Hence, from the perspective of consumers at $t = 0$, the excessive consumption is not worth the delayed costs that it brings.[7] But a consumer at time t = 0 knows that, even though he would like to moderate his consumption when $t = 1$ arrives, he might be tempted to consume in excess. Thus, he believes that his future consumption decision will be based on the following (using Eq. (1b)), and he will consume low if: $U_H - pI - \hat{\beta} C_H < U_L \Leftrightarrow U_H - pI - U_L < \hat{\beta} C_H$, else he will consume high. So, for a consumer with a perceived discount factor $\hat{\beta}$, his future (perceived) consumption in period 1 would be to consume high, if $\hat{\beta} \leq \frac{\Delta U - pI}{C_H}$ and to consume low, if $\hat{\beta} > \frac{\Delta U - pI}{C_H}$. The actual consumption in period 1 will be driven by the true β (see Eq. (1a)). So in period 1, he will consume high if $\beta \leq \frac{\Delta U - pI}{C_H}$, and he will consume low if $\beta > \frac{\Delta U - pI}{C_H}$. Thus, the consumer population can be categorized into three distinct combinations of prediction and consumption (see Fig. 2):

I. Inaccurate Hedonists (IH): $\beta C_H \leq U_H - pI - U_L < \hat{\beta} C_H \Leftrightarrow \begin{cases} \beta \leq \frac{\Delta U - pI}{C_H} \\ \hat{\beta} > \frac{\Delta U - pI}{C_H} \end{cases}$. These consumers predict to consume low but consume high.

II. *Accurate Hedonists* (AH) : $U_H - pI - U_L > \hat{\beta} C_H \Leftrightarrow \hat{\beta} \leq \frac{\Delta U - pI}{C_H}$ and $\beta \leq \hat{\beta}$. These consumers predict to consume high and consume high.

III. *Accurate Ascetics* (AA) : $U_H - pI - U_L < \beta C_H \Leftrightarrow \beta > \frac{\Delta U - pI}{C_H}$ and $\hat{\beta} > \beta$: These consumers predict to consume low and consume low.

IH consumers believe that their short-term and long-term behaviors are in harmony, but these behaviors are, in fact, in conflict. This conflict results in time-inconsistent behavior wherein these consumers expect to consume in moderation but end up consuming excessively. AH's short-term and long-term behaviors are in disharmony, but they have a better awareness of this disharmony. Finally, AA consumers predict they will consume low and in fact, do consume low. Note that within all these segments, there is considerable heterogeneity in terms of both actual self-control and beliefs about self-control.

2.2. Preventative device (PD) available

We now assume that a PD is available at a vanishingly small cost (i.e., it is available at positive but close to zero price, denoted by T).[8] Fig. 3 shows a consumer's decision process when she "buys" a PD in period 0 (at cost $T \approx 0$), and removes her short-term danger associated with excessive consumption in period 1. Note that in period 1, a consumer *expects* to consume high with PD if $\hat{\beta} < \frac{\Delta U}{C_H}$, whereas without a PD, she expects to consume high if $\hat{\beta} < \frac{\Delta U - pI}{C_H}$. On the other hand, a consumer's *actual* consumption

[6] The probability of short-term danger p, and the harm this danger brings (I) could be combined into a single short-term loss (L). However, we keep these separate for the easy interpretation in term of p.

[7] Our assumption of $\Delta U > pI$ and $\Delta U < C_H$ implies that there is a "conflict" between the "short-term self" against the "long-term self": This is the fundamental tension that has been highlighted in the self-regulation literature. Also if $\Delta U < p^*I$, then everybody in the population will have moderate consumption, and there will be no need for preventative (commitment) devices.

[8] We later relax this assumption and analyze the endogenous price of the device.

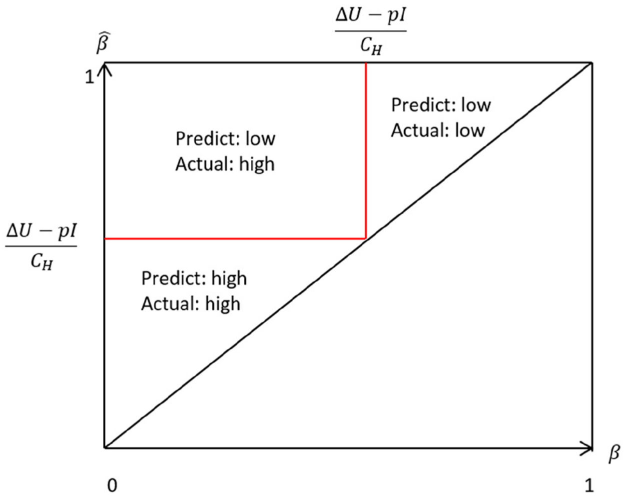

Fig. 2. Predicted and actual consumption with no preventative device.

with PD is high if $\beta < \frac{\Delta U}{C_H}$ and her actual consumption without a PD is high if $\beta < \frac{\Delta U - pI}{C_H}$. When consumers are highly overconfident (a large $(\hat{\beta} - \beta)$), then expectations and actual behavior will diverge. Consumers' ex-ante predictions and ex-post-outcomes in the presence of a PD can be put into one of six regions. These regions are shown in Fig. 4 and described below.

2.3. Low self-control: regions 1, 2 & 3

2.3.1. Region 1 (R1): $\hat{\beta} > \frac{\Delta U}{C_H}, \beta < \frac{\Delta U - pI}{C_H}$

These consumers are IHs with varying levels of overconfidence. Without a PD, a consumer's ex-ante prediction is moderate consumption that results in utility: $U_0(L) = \beta U_L$. Her actual ex-post-consumption is high with utility: $U_0{}'(H) = \beta(U_H - pI) - \beta C_H$. On the other hand, with a PD, her ex-ante prediction is moderate consumption with utility $U_0(L) = \beta U_L - T$, while her actual ex-post-consumption is high with utility $U_0{}'(H) = \beta U_H - \beta C_H - T$. A comparison of the above calculations shows that, even if a PD is available at a vanishingly small (but positive) T, consumers in R1 will choose not to use the PD because of high overconfidence (a large $(\hat{\beta} - \beta)$). Note that because $T \approx 0$, $\beta(U_H - pI) - \beta C_H < \beta U_H - \beta C_H - T$, these consumers would have been ex-post-strictly better off using the PD that they eschew.[9] To use our earlier example of drunk driving, these consumers feel confident that they will drink moderately and thus feel that they do not need a designated driver; however, they end up drinking heavily and putting themselves in harm's way. They would have benefited from a net, but they choose not to have one, and they fall into harm. Consumers' ex-ante predicted consumptions with and without the use of PD, and whether they decide to use a PD, can similarly be derived for the other regions. These derivations are presented in Table 2. We briefly discuss these below.

2.3.2. Region 2 (R2): $\frac{\Delta U - pI}{C_H} < \hat{\beta} < \frac{\Delta U}{C_H}, 0 < \beta < \frac{\Delta U - pI}{C_H}$

These consumers are also IHs but with lower overconfidence than the ones in R1. A comparison of predicted outcomes with and without a PD in Table 2 shows that no PD will be used if $T > \hat{\beta}(\Delta U - C_H)$. This condition always holds because RHS is negative as per Eq. (3), whereas LHS is positive; thus, IH consumers in R2 also end up not using a PD. As with R1 consumers, the consumers in this region end up consuming high when they would have been strictly better off using a PD since it is available at $T \approx 0$. Note that, like consumers in R1, consumers in R2 also have a low level of self-control but a moderate level of overconfidence. Specifically, consumers in R1 do not use a PD because they feel that they do not need it, whereas R2 consumers do not use PD because *they perceive that, if they use a PD, they will not have enough discipline to consume in moderation because of the removal of the short-term danger*. In other words, they try to use short-term danger as a regulation device. To return to our alcohol overconsumption example, these consumers feel they will moderate their drinking when they do not have a designated driver. The eschewing of the PD acts as a commitment device for them, but this decision fails them. Metaphorically, these consumers would have been helped by a net, but they do not use one, to help them self-regulate. They end up falling with no net to catch them.

[9] Please note that a consumer's prediction is based on $\hat{\beta}$ while the consumption is based on β.

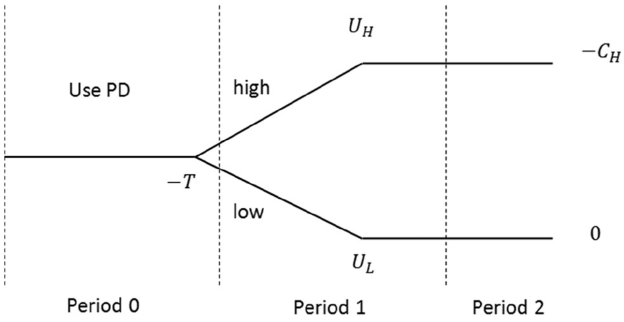

Fig. 3. Timing with a preventative device.

2.3.3. Region 3 (R3): $\beta<\hat{\beta}<\frac{\Delta U-pI}{C_H}$, $0<\beta<\frac{\Delta U-pI}{C_H}$

The consumers in this region are AHs who are sophisticated about their fallibility. A comparison of predicted outcomes with and without a PD in Table 2 shows that they will use a PD. R3 consumers are highly impatient, but they also have an awareness of their impatience (low $(\hat{\beta}-\beta)$), so they end up using a PD, and they are better off for having used the PD. In the example of drunk driving, R3 consumers know that they will be drinking heavily and are wise enough to take a designated driver with them. In other words, AH consumers in R3 carry a net and even though they fall and are saved by the net.

2.4. Moderate self-control (regions 4 & 5)

2.4.1. Region 4 (R4): $\frac{\Delta U}{C_H}<\hat{\beta}$, $\frac{\Delta U-pI}{C_H}<\beta<\frac{\Delta U}{C_H}$

These consumers are sophisticated, with a moderate discrepancy between their desired and actual behavior and fall into the earlier identified category of AAs. Comparing predicted outcomes with and without a PD in Table 2, it is clear R4 consumers will not use a PD. Consumers in R4 think that they have a high ability to self-regulate (high $\hat{\beta}$), although they have a moderate level of actual self-control (moderate β). These consumers have a moderate level of discrepancy between their perceived and true selves ((a moderate $(\hat{\beta}-\beta)$) and they end up with moderate consumption and are better off not using a PD. Interestingly, the reason these consumers eschew a PD is because they want to save on the (minuscule) PD cost of T. But there is a twist in their outcome: these consumers' eschewing of a PD results not just in savings of T, but *also in moderation of their consumption*. In the driving example, the consumers do not want the (small) hassle of finding a designated driver because they feel that they will be moderate in their drinking; however, *they end up being moderate in drinking because they do not have a driver with them*. They are regulated in the way that R2 hoped to be by the looking short-term danger. In terms of our earlier terminology, these AAs do not pay for a net because they feel that they do not need it, and they end up not falling *because of* the danger associated with the absence of net.

2.4.2. Region 5 (R5): $\beta<\hat{\beta}<\frac{\Delta U}{C_H}$, $\frac{\Delta U-pI}{C_H}<\beta<\frac{\Delta U}{C_H}$

These consumers are also AAs but with somewhat higher sophistication compared to R4 consumers. R5 consumers also would not use a PD. These consumers have moderate self-control problems and also have a moderate level of perceived self-control and

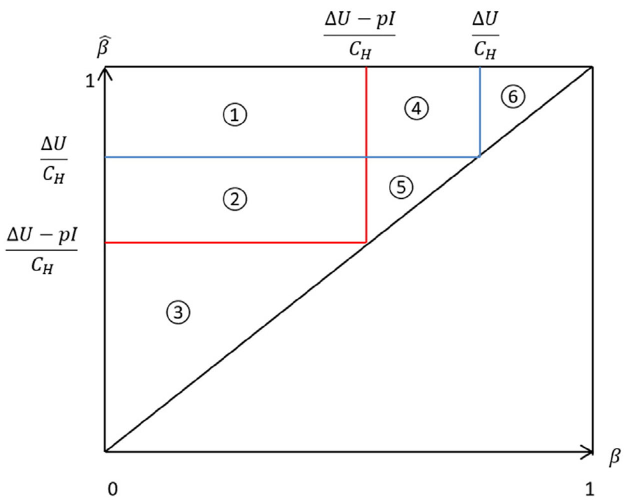

Fig. 4. Segments with a preventative device.

Table 2
Consumer decision making with and without a preventative device.

Region	*No PD available*		*PD available*		Decision
	Prediction	Actual	Prediction	Actual	
R1	$U_0(L) = \beta U_L$	$U_0{}'(H) = \beta(U_H - pI) - \beta C_H$	$U_0(L) = \beta U_L - T$	$U_0{}'(H) = \beta U_H - \beta C_H - T$	No PD
R2	$U_0(L) = \beta U_L$	$U_0{}'(H) = \beta(U_H - pI) - \beta C_H$	$U_0(H) = \beta U_H - \beta C_H - T$	$U'_0(H) = \beta U_H - \beta C_H - T$	No PD
R3	$U_0(H) = \beta(U_H - pI) - \beta C_H$	$U_0{}'(H) = \beta(U_H - pI) - \beta C_H$	$U_0(H) = \beta U_H - \beta C_H - T$	$U_0{}'(H) = \beta U_H - \beta C_H - T$	Use PD
R4	$U_0(L) = \beta U_L$	$U_0{}'(L) = \beta U_L$	$U_0(L) = \beta U_L - T$	$U_0{}'(H) = \beta U_H - \beta C_H - T$	No PD
R5	$U_0(L) = \beta U_L$	$U_0{}'(L) = \beta U_L$	$U_0(H) = \beta U_H - \beta C_H - T$	$U_0{}'(H) = \beta U_H - \beta C_H - T$	No PD
R6	$U_0(L) = \beta U_L$	$U_0{}'(L) = \beta U_L$	$U_0(L) = \beta U_L - T$	$U_0{}'(L) = \beta U_L - T$	No PD

Note: The regions are defined in Section 2.2 in the paper. Note that a consumer arrives at the predicted behavior using $\hat{\beta}$: her perceived self-control parameter. The utility expression involves β- the actual self-control parameter. For example, a consumer in R1 predicts low consumption because $U_L > U_H - pI - \hat{\beta} C_H$ while they end up consuming high because $U_L < U_H - pI - \beta C_H$.

thus a low level of overconfidence (a low $(\hat{\beta} - \beta)$) and these consumers correctly follow using “no PD” as a strategy to avoid overconsumption. Also, comparing with R2 (who are IHs), although both these sets of consumers have similar levels of perceived self-control, R5 consumers end up with moderate consumption because they also have a moderate level of actual self-control. In our driving example, these consumers are aware that they would be tempted to drink heavily if they had a designated driver; hence, they go to the party without one. In this case, the strategic use of eschewing a PD counterintuitively leads to self-regulation, making R5 consumers better off. These consumers fly without a net because they realize that having a net would lead them to fall, and they are right.

2.5. High self-control (region 6)

2.5.1. Region 6 (R6): $\frac{\Delta U}{C_H} < \beta < \hat{\beta} < 1$

These consumers do not have self-control problems and thus are the most accurate as well as most self-regulated among AAs. R6 consumers correctly perceive a high level of self-control (a very low $(\hat{\beta} - \beta)$). They end up eschewing a PD and consumer in moderation. These consumers do not fall whether there is a net or not, and since a net has a small cost, they are correct to eschew it.

To summarize, our model allows us to make numerous predictions about consumer adoption of, and need of, a PD that go far beyond just whether people will use one or not. We wish to point out that the rich behavior exhibited by consumers in our model stems from the heterogeneity in both the self-control and perception of self-control parameters. Again, this is an important contribution of our model (and a departure from classic models and previous research such as Jain and Li (2018)[10]). Specifically, we show: (1) Sophisticated AHs use a PD if they have a high impatience (R3) while most patient AAs avoid its use since they do not need it (R6); (2) Naïve IHs avoid a PD and end up potentially hurting themselves (R1 and R2). (3) Lastly, we showed that sometimes, “flying without a net” has a positive outcome. In fact, not using the PD acts as a tool for moderating these consumers' consumption, either strategically (R5), or accidentally (R4). Table 3 summarizes this discussion.

3. Analysis and results

In this section, we use our framework and results to examine the welfare implications of public policies and marketing interventions often used in the context of vice good consumption. Specifically, we examine the effectiveness of (1) uniform enforcement of the use of a PD, and (2) raising the perceived probability of the short-term danger of the consumption of vice goods. Also, there has recently been a surge in the interest in market-based solutions to individuals' self-control problems (Brocas, Carrillo, & Dewatripont, 2004b) and in that tradition, we also examine the case in which a monopolist seller supplies a PD, and we characterize the optimal pricing of the provision of such a PD. While performing welfare analyses, we also account for the fact that the short-term effects of excessive vice consumption also have significant negative externalities in addition to the short-term danger. For example, drunk driving carries an external cost to society as others could get physically hurt or have their property damaged. Because policymakers might focus either on externalities like additional drunk driving accidents or violence emanating from drug use or on consumer welfare, or both, our welfare calculations account for both consumer welfare as well as the effect on others via an externality.[11]

[10] If people had self-control problems but were sophisticated about it (meaning, $\beta = \hat{\beta}$) then in Fig. 4, we would only have two segments (with all the consumers massed on the diagonal line) with correct anticipations (“Predict High/Consume High” or “Predict Low/Consume Low”). This lack of overconfidence would lead to consumers always correctly using or not using a PD, a result that does well reflect the empirical reality of people often being mistaken about their self-regulation.

[11] Gruber and Kőszegi (2004) report that estimated externality the cost of a pack of cigarettes based on prior empirical work to be around *40 cents per pack* (pp. 1961). Their calibrations of a β/δ model using the *Consumer Expenditure Survey* reveals that a pack of cigarette costs about *$35.64 in terms of loss of life expectancy*. In other words, their estimate of costs of “internalities” is about *100 times more than the cost of externalities*. Going beyond these numbers, we feel that it is important to consider both the externalities as well as consumer welfare (“internalities”) to analyze the trade-offs among different policies. We are grateful to an anonymous reviewer for bringing forth this point.

Table 3
Summary of consumer decisions and intuitive reasoning.

Segment	PD?	Why?	Result?
R1: IH	No	Don't realize they need it	Fall with no net
R2. IH	No	Trying to increase risk (and thus self-control)	Fall with no net
R3. AH	Yes	Realize they need it	Fall with a net
R4. AA	No	Cost of PD	Do not fall (because no net)
R5. AA	No	Incentive of having no net not to fall	Do not fall (because no net)
R6. AA	No	Not tempted	Do not fall whether there is a net or not

3.1. Policy Intervention 1 - uniform enforcement of preventative devices

Policymakers often advocate the uniform enforcement of PDs to remove the immediate dangers facing individuals consuming vice goods. For example, there have been extensive campaigns promoting carrying condoms during nights out (BBC News, 2006; Telegraph, 2009), legislative attempts at making it compulsory for large event organizers to provide designated drivers for participants (May 11, Las Vegas Sun, 2003), and initiatives to hand out fresh needles for use with illegal drugs (January 17, *Baltimore Sun*, 2014). Using our framework, we can analyze the welfare impact of a policy that makes it compulsory to the consumers of vice goods to use PDs.

Fig. 5 illustrates the resulting consumer behavior when the use of a PD is made compulsory. In the base model, the subset of consumers who do not voluntarily adopt a PD (R1, R2, and R4, R5, R6) are now forced to adopt it. In reality, even when a policy provision makes something compulsory, compliance is likely to be imperfect, so our model also includes a parameter τ that represents the probability of adoption by segments which are forced to adopt as a consequence of policy. Thus $(1-\tau)$ represents the non-compliance under the policy. As before, we assume that PD is made available without any cost (i.e., $T \approx 0$). Note that consumers in regions R3 and R6 do not see any change in welfare (the former segment adopts PD under both the regimes while the latter consume in moderation whether they have a PD or not). Therefore, the uniform enforcement affects regions R1 & R2 and R4 & R5 (compare Figs. 4 and 5). Specifically, in R1 & R2 the policy now prevents mistaken rejection of PDs and thus *makes the consumers who comply* better off, as the policy was aiming to do. On the other hand, in R4 & R5, wherein consumers were able to moderate their consumption by using the immediate danger as a tool to moderate excessive consumption, a policy compliance results in over-consumption because of the removal of the short-term danger via the PD and hence counterintuitively makes them *worse off than they would be without the compulsory PD.*

The overall welfare effect of making the PD compulsory thus depends on the size of the relative welfare of the change in the "winning" segments (R1 & R2) versus the "losing" segments (R4 & R5) under a uniform provision of PDs. To quantify this effect, for the rest of the paper, we assume the consumer population to be uniformly dispersed in the upper triangle as in Fig. 2. More formally, the joint distribution of β and $\hat{\beta}$ is uniform and the joint density function is: $f_{\beta\hat{\beta}} = \begin{cases} 1 \text{ } if \text{ } 0<\beta<\hat{\beta} \\ 0 \text{ } otherwise \end{cases}$, and $\int_0^1 \int_0^{\hat{\beta}} f_{\beta\hat{\beta}} d\beta d\hat{\beta} = \frac{1}{2}$.

There are two additional layers to this welfare analysis. First, it is important to note that welfare analysis for time-inconsistent agents has included a discussion of which type of welfare is more relevant. If an individual is modeled as having "multiple-selves," then a natural question arises from a social marketing or policy perspective: which self's welfare is paramount (Bernheim & Rangel, 2009; Camerer, Issacharoff, Loewenstein, O'Donoghue, & Rabin, 2003)? In our terminology, is the $t = 0$ (long-term), or

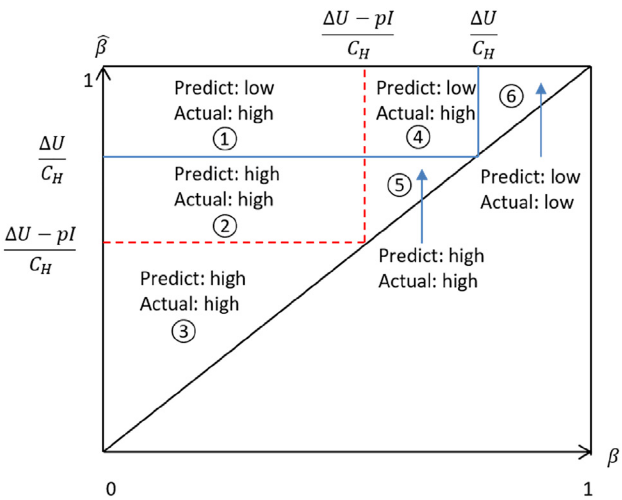

Fig. 5. Predicted and actual consumption with uniform enforcement of PD.

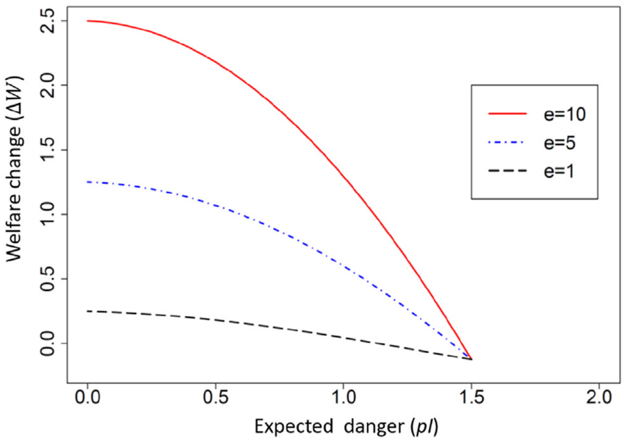

Fig. 6. Welfare change by expected short-term danger (pl) and externality level.

$t = 1$ (short-term) utility the correct one to focus on for social marketing and social welfare purposes? By far the most widely used criterion is to consider the welfare of the long-term perspective of an agent (O'Donoghue & Rabin, 2000), under the reasonable assumption that the forward-looking agent unswayed by temptations has a better perspective and is driven less by short-term visceral responses. This approach thus requires the construction of welfare measures from the long-term outlook of the agent- an approach that we use below as we fully analyze welfare under a compulsory PD. Second, as mentioned, when conducting a policy evaluation in addition to a policy's effect on the consumer, we need to account for the externalities associated with excessive consumption absent PD. For example, a drunk driver on the road not only puts himself in danger but also puts the lives of other drivers and pedestrians in danger as well. Let $e \geq 0$ denote the expected externality cost incurred upon the society by a consumer putting himself into danger. When the uniform enforcement policy is not in place, the externalities originate from R1 and R2 and are denoted by: $E_{NE} = e\frac{\Delta U - pI}{C_H}(1 - \frac{\Delta U - pI}{C_H})$. When the policy is in place, the externality costs emanate from segments R1, R2 (those who do not comply). Note that when a compulsory PD is in place, no externalities are emanating from segments R4 and R5, whether they comply or not. Those who do not comply moderate their consumption, so there are no long-term or short-term costs and no externalities. Those who do comply, consume in excess, so there are long-term costs, but since they comply (voluntarily) with the PD, there are no short-term costs and hence no externalities associated with them. For a compliance rate of τ, the total negative externalities under the enforcement policy are: $E_{UE} = e[(1-\tau)\frac{\Delta U - pI}{C_H}(1 - \frac{\Delta U - pI}{C_H})]$.

To summarize, the compulsory provision of PD results in the following welfare changes: (1) R1 & R2: Compliant consumers among these consumers end up removing short-term danger via PD provision (*positive effect*), (2) R4 & R5: Compliant consumers among these consumers end up consuming in excess and incurring long-term costs (*negative effect*). Our focus welfare analysis focuses upon the *overall change* in the welfare as a consequence of the policy, and the algebraic expression for this net effect is presented as ΔW in the Technical appendix (under the proof of Prop. 1). When ΔW is positive, the policy in beneficial while it is welfare-reducing otherwise. An analysis of the algebraic expression of ΔW can be used to derive Proposition 1.[12]

Proposition 1. Under the compulsory provision of a preventative device, there always exists a p^*, such that overall welfare is <u>lower</u> when $p \geq p^*$ and <u>higher</u> when $p < p^*$ (relative to voluntary adoption of a preventative device).

The above results suggest that overall welfare could increase under the compulsory provision of PDs. Still, the gains obtained by a subset of consumers would always come at the cost of others. In other words, the enforcement of PDs is not welfare improving for everyone. It is important to point out that policymakers are likely to be especially prone to enforce compulsory PD's when the short-term danger is high, such as on campuses where excess drinking is prevalent. However, our model shows that the overall welfare improvement is least likely when the risk of short-term danger is high. We discuss this intuition below.

When the probability of short-term danger is low, fewer people will use the danger as a credible deterrent to self-regulate. Consumers who are more likely to eschew a PD (under the low probability of danger) to self-regulate are likely to be the ones who are most (incorrectly) overconfident, and these people (R1 and R2) will be precisely the ones to benefit the most from the compulsory provision of a PD. On the other hand, when the probability of the danger is high, many consumers- overconfident (R1 and R2) and relatively sophisticated (R3 and R4) avoid a PD to self-regulate. The former types are avoiding the PD mistakenly while the latter types are avoiding it correctly, and the fraction of correct users (relatively) increases as the danger becomes more

[12] To keep our analysis tractable, we focus on the probability of short-term danger while fixing other variables, whenever possible. The qualitative results from our model should hold in general, though. Also note that P1 is a statement of sign of *welfare change* as a result of policy and not about the monotonicity, that is $d(\Delta W)/dp$.

likely. Under these conditions, compulsory provision is going to be hurtful on average as it is likely to push more such consumers into overconsumption (see Fig. 5). Hence we obtain the counterintuitive outcome that at the high levels of the danger, the overall welfare effect of the policy is likely to be negative. Also, note that the consumers with high sophistication are unaffected by the compulsory provision of the PD: the least patient among the sophisticated would overconsume (R3) and would have purchased PD anyway; while the most patient (R6) will consume in moderation whether they have a PD or not.

A corollary of P1 is that at the lower levels of risks, PD enforcement is welfare improving, and the greater the welfare improvement, the greater the level of externality (e). On the other hand, in the zone of higher risk where PD enforcement is welfare decreasing, the sharper is the welfare reduction, the higher the level of negative externality (see Fig. 6). In other words, enforcement is really effective *when the short-term danger is low-to-moderate, and the externality is high*. On the other hand, quite counterintuitively, the implementation of a PD would be of little use when both the *short-term threat and the externality are high*. The latter conditions are precisely the conditions often advocated for the provision of PDs. Another surprising implication is that welfare change is non-monotonic in the compliance rate: compliance rate and degree of danger do not have straightforward effects on the success of this policy. Specifically, when the probability of danger is relatively low (high), the provision of PD is welfare-enhancing (reducing), and an increase in the compliance rate further improves (reduces) the welfare.[13]

3.2. Policy Intervention 2 - increasing the perceived probability of danger ($\hat{p}>p$)

We now turn to another commonplace policy intervention: highlighting short-term dangers via advertising, PSA's, and social marketing initiatives aimed at deterring the excessive consumption of vice goods. In terms of our model, these messaging strategies can be succinctly summarized as efforts intended to raise the perceived probability of short-term danger to $\hat{p}$ higher than the actual probability, p. A recent well-publicized example is the “Think!” video release campaign sponsored by the UK's Department for Transport, highlighting the dangers of drunk driving. One video shows men standing in front of the mirror in a pub bathroom washing their hands. Suddenly, a loud crash sends each prank victim leaping backward as a woman's bloody face appears from the other side of the broken mirror (Telegraph, 2015). Similar fear-inducing campaigns have run in many other parts of the world (Castillo-Manzano, Castro-Nuño, & Pedregal, 2012) and many different contexts (New York Times, 2011), including anti-drug campaigns (Morgan, Palmgreen, Stephenson, Hoyle, & Lorch, 2003) and anti-AIDS advertising (Dillard, Plotnick, Godbold, Freimuth, & Edgar, 1996). Our model allows us to study the welfare consequences of the efforts of social marketing campaigns that succeed in raising individuals' perceived probability of the short-term dangers associated with vice good consumption.

In terms of our model, the increase in perceived probability has two effects: (1) it expands R2 (recall that these are the individuals who mistakenly do not use the PD to self-regulate) by shrinking R3, and, (2) it expands R4 and R5 (wherein individuals are able to exercise self-control) at the expense of R2 and R3 (see Fig. 7). The former effect is captured by region S1, and the latter by region S2, respectively. Let $\hat{p}=p+\Delta p$, $\Delta p>0$, then:

$$S_1=\frac{\Delta pI}{C_H}\cdot\frac{\Delta U-pI-\Delta pI}{C_H}$$

$$S_2=\frac{\Delta pI}{C_H}\frac{1}{2}\left[\left(1-\frac{\Delta U-pI-\Delta pI}{C_H}\right)+\left(1-\frac{\Delta U-pI}{C_H}\right)\right]=\frac{\Delta pI}{C_H}\left(1-\frac{\Delta U-pI}{C_H}+\frac{\Delta pI}{2C_H}\right).$$

Again, we define the change in welfare as $\Delta S=S_2-S_1$. The former effect (expansion of S_1) is welfare decreasing, whereas the latter effect (expansion of S_2) is welfare enhancing. Again, we can see that the impact of the policy of raising the perceived probability is not Pareto improving because it expands S_2 at the cost of S_1. Specifically, we obtain the following result[14]:

Proposition 2. When the perceived probability of the short-term danger is raised by Δp through PSAs, the following hold:

- **(i)** *For* $2(\Delta U-pI)-C_H>0$*: If* $\Delta pI\geq\frac{2[2(\Delta U-pI)-C_H]}{3}$*, welfare is higher, i.e.,* $\Delta S>0$*, and* ΔS *increases with* Δp*. If* $\Delta pI<\frac{2[2(\Delta U-pI)-C_H]}{3}$*, welfare is lower, i.e.,* $\Delta S<0$*, and* ΔS *decreases with* Δp *when* $\Delta pI<\frac{2(\Delta U-pI)-C_H}{3}$ *and increases with* Δp *when* $\Delta pI\geq\frac{2(\Delta U-pI)-C_H}{3}$.
- **(ii)** *For* $2(\Delta U-pI)-C_H\leq 0$*, welfare is always higher: “winning” segments are always greater than “losing” segments, i.e.,* $\Delta S>0$ *and* ΔS *increases with* Δp.

Condition (i) implies that the temptation is relatively high, as compared with the long-term costs, and increasing the perceived probability of danger largely affects only the relatively unsophisticated consumers (R1 and R2). The proportion of the increase in the number of consumers in R1 and R2 exceeds the increase in the proportion of consumers who use the immediate danger to truly moderate consumption (R3 and R4), except when the short-term danger is sufficiently high. Intuitively, a (small) increase

[13] The mathematical details of these supplementary analyses are available upon request from authors.
[14] We assume that I is sufficiently large so that Δp is always between 0 and $1-p$.

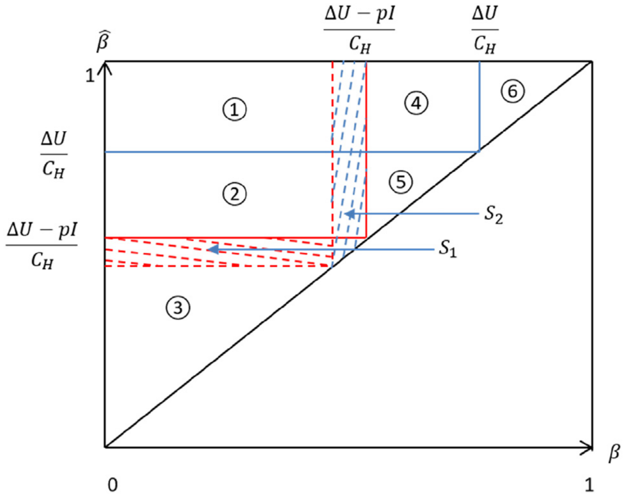

Fig. 7. Change in segment sizes with increased in perceived short-term danger ($\hat{p}>p$).

in the perceived probability of the short-term danger will worsen things and will induce more unsophisticated consumers into believing that a threat of danger will make them moderate their consumption (see Fig. A2). Condition (ii) implies that when the threat of temptation is relatively low, the sophisticates are more likely to use the increased perceived probability of danger as a tool to moderate consumption. Still, as the perceived probability goes up, more of relatively unsophisticated consumers will start to use the increased perceived probability as a tool as well (see Fig. A3).

In summary, although the perceived probability increase is not Pareto improving because it always makes some consumers worse off, at the lower levels of temptation, the increase in perceived probability induced by the campaign can make more consumers better off in aggregate. The most interesting implication of P2 is that, in the face of high temptations, aggressive messages, including strategies such as shock value, might work. On the other hand, strategies that enhance the danger component only moderately might be overall counterproductive and even increase negative behaviors.[15]

4. Extensions

In this section, we present three extensions of our primary model. We have so far abstracted away from the price of the vice good and assumed it as part of the utility from consumption. However, pricing could help marketers and public policymakers improve welfare. In the first extension, we briefly discuss the implications of the pricing of the vice good in our model, because the pricing of vice goods is a common way to conceptualize policies for self-control (O'Donoghue & Rabin, 2006). In the second extension, we endogenize the pricing policy of a monopolist selling a PD. Finally, after considering pricing policies we analyze the impact of an intrapersonal strategy on the part of consumers of becoming strategically pessimistic about overcoming the self-regulation problems arising from overconfidence.

4.1. Implications of pricing of vice goods

The main aim of this extension is to provide an intuition for pricing incentives for the seller of a vice good, and then contrast it with the incentives that a policymaker might have to intervene for welfare purposes. As such, we keep this extension simple and leave the complete characterization of optimal pricing in this setting for future research.

Let q be the price of excessive consumption and normalize the price of moderate consumption to zero. There are two cases to consider:

Case A. If q is low such that $\Delta U - pI - q \geq 0$, the regions are similar to those in Fig. 4, and the size of these regions now depends upon the q, in addition to other parameters, as seen in Fig. 8. Specifically, note that:

i. The size of R1 is $\frac{\Delta U - pI - q}{C_H} \cdot (1 - \frac{\Delta U - q}{C_H})$ which changes non-monotonically with the price of the excessive consumption, i.e., it increases in q when $0 \leq q \leq \Delta U - \frac{1}{2}pI - \frac{1}{2}C_H$ and decreases in q when $\Delta U - \frac{1}{2}pI - \frac{1}{2}C_H \leq q \leq \Delta U - pI$.

ii. The size of R2 is $\frac{\Delta U - pI - q}{C_H} \cdot \frac{pI}{C_H}$ which decreases monotonically in q.

[15] We also carried out a more comprehensive welfare analysis in which we take into account the number of consumers affected, as well as the change in their utilities. The results are very similar.

iii. The size of R3, $\frac{1}{2}(\frac{\Delta U-pI-q}{C_H})^2$, monotonically decreases in q.

iv. The size of R4, $\frac{pI}{C_H}\cdot(1-\frac{\Delta U-q}{C_H})$ monotonically increases in q.

v. The size of R5, $\frac{pI}{C_H}\cdot\frac{pI}{C_H}$, does not change in q.

vi. Finally, the size of R6, $\frac{1}{2}(1-\frac{\Delta U-q}{C_H})^2$ increases in q.

Case B. If q is high, i.e., $\Delta U - pI < q < \Delta U$, R1, and R4 will merge into a single region with a size $\frac{\Delta U-q}{C_H}\cdot(1-\frac{\Delta U-q}{C_H})$. Let us call this region R14. The regions R2, R3, and R5, will merge into a region with a size $\frac{1}{2}(\frac{\Delta U-q}{C_H})^2$. Let us call this region R235.

i. The region R14 first increases in q then decreases in q.

ii. The size of R235 monotonically decreases in q.

iii. The size of R6 still has a size of $\frac{1}{2}(1-\frac{\Delta U-q}{C_H})^2$, which increases in q.

4.2. Profit implications of vice good pricing

In our simple setting, the vice-good seller can earn revenues by selling only to consumers in R1, R2, and R3 (the hedonists who are tempted) because consumers in R4, R5, and R6 moderate their consumption. When the price q is high ($\Delta U - pI < q < \Delta U$), R1, R2, and R3 will disappear. Therefore, the firm will only charge a lower price (($q < \Delta U - pI$) to maximize revenue. Then the firm's revenue from selling a vice good is the price times the demand, i.e.:

$$\Pi = q\left(\frac{\Delta U-pI-q}{C_H}\cdot\left(1-\frac{\Delta U-q}{C_H}\right)+\frac{\Delta U-pI-q}{C_H}\cdot\frac{pI}{C_H}+\frac{1}{2}\left(\frac{\Delta U-pI-q}{C_H}\right)^2\right)$$
$$= q\left(\frac{\Delta U-pI-q}{C_H}\cdot\left(1-\frac{1}{2}\left(\frac{\Delta U-pI-q}{C_H}\right)\right)\right) = q\left(\frac{\Delta U-pI-q}{C_H}\right)\cdot\left(1-\frac{1}{2}\left(\frac{\Delta U-pI-q}{C_H}\right)\right)$$
$$= -\frac{1}{2C_H^2}q(\,q-(\Delta U-pI))\cdot(q-(\Delta U-pI-2C_H))$$

Quite intuitively, pricing would reflect a trade-off between the revenue from each individual and the total demand from R1, R2, and R3 (as seen in Fig. 9). However, note that from a policy perspective, the regulators wish to either moderate consumption of R1, R2, and R3 or if not, use PDs to save these consumers from short-term dangers. A revenue-maximizing firm does not have an incentive to make people moderate their consumption (via high q) by making R1, R2, and R3 disappear.[16] The presence of these regions (especially R1 and R2, who are inaccurate and hedonists) could motivate policymakers to provide regulations like the provision of PDs, as the danger associated with excessive consumption is likely to be in place even in the face of endogenous prices.

4.3. Market provision of PD at endogenous price (T)

Because the consumption of many vice goods can harm consumers, and as we have shown that commonplace policy interventions might be imperfect, markets could potentially play a role in the provision of PDs. Examples of such price-based PDs include condoms (to prevent STDs), designated drivers provided at parties and bars (to prevent road accidents), radar detectors (to avoid speeding tickets), and clean needles (to prevent HIV infection). In this section, we derive the price charged for PDs by a monopolist seller; we endogenize the cost of PD (T) considered so far in the paper.[17] The characterization of this pricing policy is important from the obvious managerial perspective for the sellers of PDs. It is also useful for policymakers and social marketers who often subsidize the provision of such PDs (e.g., in many parts of Europe the clean needles used by drug users come from the local city authorities). Assuming the marginal cost to be zero, from the perspective of a PD seller, she would set up T to maximize the expected revenue Π. Note that only consumers in R3 (Fig. 4) will consider buying a PD.

[16] As a reminder, the revenue calculations do not include the revenue from moderate consumption that might be substantial in many contexts. See Jain and Li (2018) for a more complete characterization of pricing in a similar context.

[17] Note that this is the pricing of PD in contrast to the pricing of vice good itself that was just discussed.

Hence, the expected revenue is given by[18]:

$$\Pi = T \cdot \; Pr\left(\frac{T}{pI}<\hat{\beta}<\frac{\Delta U - pI}{C_H}\right) = T \cdot \frac{\frac{1}{2}\left(\frac{\Delta U - pI}{C_H}+\frac{T}{pI}\right)\left(\frac{\Delta U - pI}{C_H}-\frac{T}{pI}\right)}{\frac{1}{2}=\left(\frac{\Delta U - pI}{C_H}\right)^2 T-\frac{T^3}{(pI)^2}} \tag{9}$$

Expression (9) can easily be maximized with respect to *T*, and we obtain the following characterization of the optimal T^*:

Proposition 3. The optimal price T^* of a preventative device is concave in probability p, i.e. T^* increases in p when $pI<\frac{\Delta U}{2}$ and decreases in p when $pI\geq\frac{\Delta U}{2}$.

This result is somewhat surprising because the very purpose of a PD is to remove the danger associated with the excessive vice good consumption; consequently, we might expect the price to increase with the greater (expected) danger. This result obtains because the price response of a PD seller to the increase in the probability of the short-term danger depends on two effects. The first effect is a standard *quality effect* by which a device with a higher pay-off[19] is priced higher in the market, while an opposing *demand effect* whereby the danger associated with excessive consumption acts as a soft commitment device by consumers (R1 & R2 and R4 & R5 respectively in Fig. 5). The perceived efficacy of *eschewing a PD* as a tool to moderate consumption increases in the probability of short-term danger, which shrinks the demand for a PD. For the lower levels of p, the former effects dominate the latter, and vice-versa. This result highlights that the pricing of PDs – in contrast to most products- could often be counterintuitive and might not be simply a monotonic function of "quality".[20]

4.4. Pessimism as a consumer response to self-control problems

Finally, we present a brief discussion on changes in outcomes when the population is systematically pessimistic about their self- regulation ability. The focus here is on the discussion of key insights, and the necessary mathematical details are found in the Technical appendix.

Consumers are usually modeled as being either sophisticated ($\hat{\beta}=\beta$) or naively overconfident in their abilities to self-regulate ($\hat{\beta}>\beta$). The approach we use here includes a consumer population with varying degrees of impatience and varying degrees of sophistication. This formulation is consistent with the large psychological literature that posits that consumers often fall prey to temptations and are systematically overconfident in their forecasts of the ability to self-regulate (Ainslie, 1975), which has motivated a burgeoning behavioral economics literature to model agents' perceived immediate gratification parameter as $\hat{\beta}\geq\beta$ (see DellaVigna, 2009 for a review).[21] Nevertheless, it is conceivable that, after repeated (and often futile) attempts at self-regulation, some consumers might become pessimistic about their ability to exercise self-control. In terms of our model, pessimism corresponds to $\hat{\beta}<\beta$; consumers (wrongly) believe that their immediate gratification urges are worse than they actually are. To examine this situation, we present a specialized case of our basic set-up in which the entire population is (weakly) pessimistic in their beliefs about their ability to self-regulate (see Fig. 10). Similar to the analysis done in Section 2, we can show that without access to a PD, consumers will be in one of three groups: 1) Those who expect to consume high and end up consuming high, 2) those who expect to consume low and end up consuming low, and 3) those who expect to *consume high but end up consuming low*.

Then, we look at the consumer choices in a scenario wherein a PD is made available for (almost) free, and we can divide the consumers into six distinct segments (see Fig. 11). As before, we find that relatively sophisticated consumers make correct decisions: They use a PD when it removes the short-term danger (R3), and do not use a PD when they can self-regulate without its use (R4, R5, R6). More interestingly, consumers with a high degree of self-regulation (high β) but high pessimism (low $\hat{\beta}$)—represented by R1 in Fig. 11—end up using a PD when it turns out to be of no use. If the cost of a PD is close to zero, this unnecessary cost is trivial but could be substantial if the PD were supplied by the market. More perversely, consumers with intermediate levels of self-regulation but high pessimism (R2) end up using a PD, and this leads to a *boomerang effect* whereby they end up consuming high *because of the use of PD*. In other words, a segment of highly pessimistic consumers who use a PD in the hopes of averting immediate danger ends up over-consuming precisely because a PD takes the threat of immediate danger off the table. Absent a PD; these consumers would have consumed in moderation. These last two outcomes suggest that if consumers use pessimism as a strategic response to the dangers of overconfidence, they are not always better off. This result stands in contrast to Heidhues and Kőszegi's (2010) finding in credit markets, showing that although overconfidence could hurt

[18] Note that we ignore externalities in the consumer demand formulation as a consumer is unlikely to fold back (small) magnitude of externalities in the individual purchase decision just like a small steel mill owner would not take into account the effect of its pollution on the global warming. Of course, at the aggregate level these externalities could add up and become significant.

[19] Note that higher pay-off here implies a removal of a higher (expected) danger.

[20] This result shares some similarities with Jain and Li's (2018) finding who document the non-monotonic effect of unhealthiness of a product on prices.

[21] We conducted a small meta-analysis about the estimates of β from the past studies and found that almost all the studies have found the empirical evidence for self-control in the studies of inter-temporal decision-making. As an example, in Brown, Chua, and Camerer (2009), an experimental setting that involved 30 rounds of play (with actual payoffs), 25 out of 26 subjects displayed $\beta < 1$, providing some tentative evidence that a bias for immediate consumption seems to be quite prevalent in varying degree.

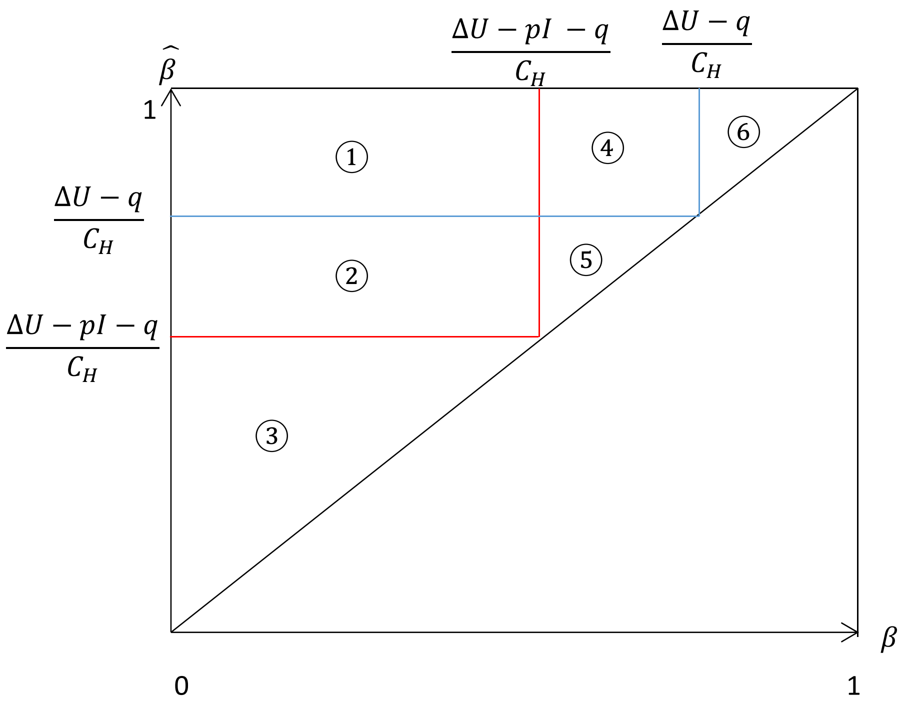

Fig. 8. Segments with a preventative device (with vice good pricing).

consumers, pessimism does not lead to any welfare loss. Of course, our context and modeling approach differs substantially from theirs; in our situation, the presence of both short and long-term risks combined with the optional PD sets up the potential for consumers to cause harm to themselves even as they try to be cautious.

5. Concluding remarks

Vices are both a heavy burden for society and a source of great attention for marketers and public policymakers. In the realm of alcohol alone, more than 10,000 people die from drunk driving in the US, and another 80,000 people die from alcohol-related causes every year (cdc.gov). Vices often carry both short-term and long-term risks, which consumers consider in their decision making, and they often consider a preventative device for the short term risk such as avoiding drunk driving using a PD. Designated drivers, ride companies specifically set up to transport inebriated customers, and taxis/rideshares are examples of these PDs. PDs are perceived as uniformly good if one takes a simple approach to the problem. Still, we contend that use of PDs is more complicated for many risks not only because there are two types of risk involved, but also because people have differing levels of temptation and that they are heterogeneously knowledgeable or sophisticated about these temptations.

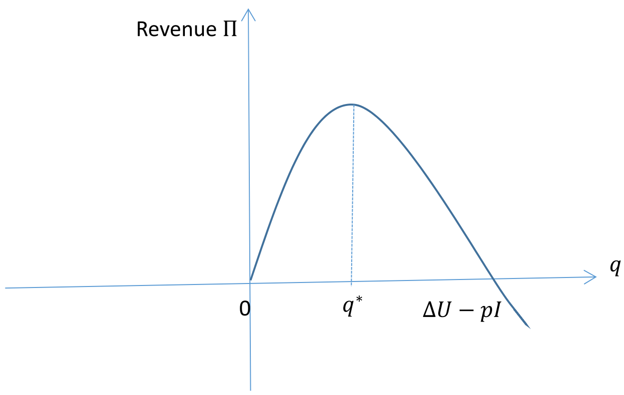

Fig. 9. Revenue maximizing price of vice good for excessive consumption.

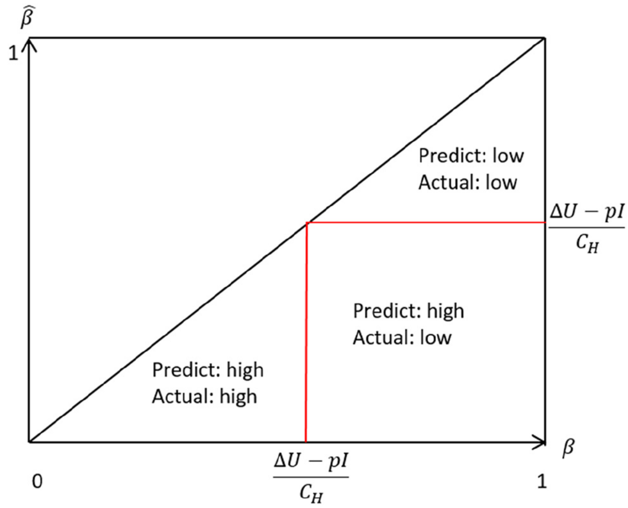

Fig. 10. Predicted and actual consumption with pessimistic beliefs ($\hat{\beta}<\beta$).

We show that PDs are not uniformly desirable, and provide a rationale that some consumers are likely to forgo the PD as a soft commitment device to keep themselves from over-consuming. For sophisticated consumers, this unusual strategy actually works. For less sophisticated consumers, the strategy can prove to be counter-productive. Our formulation is the first that we know of to suggest that consumers might try to increase commitment by forgoing safety, and we expect that future empirical research might explore this sometimes useful and sometimes disastrous consumer behavior further.

We take both public policy and a marketing approach to try to highlight some of the counterintuitive findings from our set-up. It might make sense at first glance to provide PDs within a population: to require designated drivers (perhaps by rendering automobiles inactive if the driver has had too much to drink), to require distribution of clean needles for addicts via a government program, and so on. One intriguing finding of our policy analyses is that a set of consumers who would otherwise consume in moderation end up engaging in excessive consumption because of compulsory PDs. This finding may seem odd. Still, there is empirical data that is consistent with it, suggesting that our model, which explains the finding, may provide a reasonable explanation for extant confusing results. For example, Arcidiacono, Khwaja, and Ouyang (2012), using the 1997 Longitudinal Survey of Youth data, find that increased and easier access to contraceptives such as condoms (a PD) increases the pool of sexually active teenagers who might have otherwise abstained. Peltzman (1975) in another related example, shows that compulsory seat belt laws could result in more hazardous behavior by a subset of drivers. Finally, there is a practice at some universities of providing free buses to students to transport to bars around the campus and back, popularly known as “bar buses”. Interestingly, in 2011 the University of Rochester suspended its bar bus service

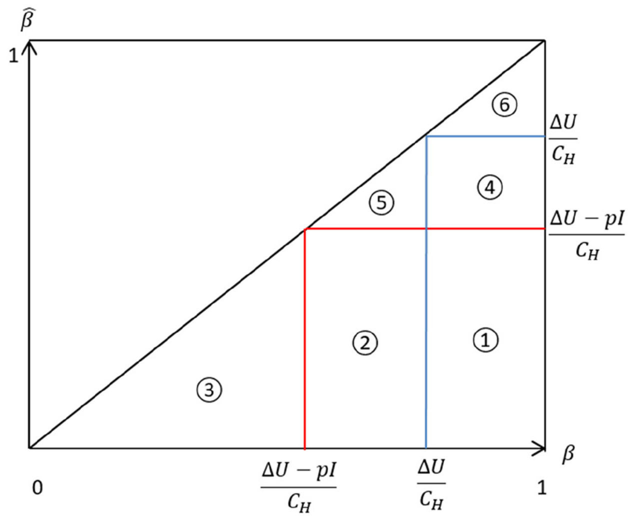

Fig. 11. Segments with a preventative device and pessimistic beliefs.

to “discourage and curb excessive drinking” – perhaps in anticipation of our finding that compulsory use of PDs can lead to excessive vice consumption (Campus Times, 2013). We are not suggesting that compulsory PDs are not a good idea, but simply pointing out a wrinkle to them that standard models could not uncover. The degree to which they encourage excess consumption depends on the size of each of our segments in the population.

Also, our study theoretically complements surging interest in empirical research on the efficacy of “counter-marketing” that seeks to reduce the consumption of vice goods (Wang, Lewis, & Singh, 2016). We show that policy interventions such as the use of PSAs to artificially inflate the short-term danger often produce unambiguously decreasing welfare. Perhaps our results help explain why PSAs have, in general, had mixed effects (Dillard et al., 1996).

Even personal strategies such as strategic pessimism might not improve consumer welfare because of the unique way that PDs interact with short-term and long-term risks. In fact, pessimism can result in excessive consumption precisely because of the unnecessary use of a PD that the pessimism can inspire. These findings suggest that perhaps a better policy that could be welfare-enhancing for all consumers is “debiasing” consumers (Gabaix & Laibson, 2006), creating (correct) consumer awareness of their self-control problems, so they understand their frailties better and become more sophisticated in the use of PDs and commitment devices. Recent push towards policies such as credit counseling and improving financial literacy in credit markets could reflect this type of intervention (Meier & Sprenger, 2013).

Our model has the potential to be extended in multiple directions. For example, the market experience could make consumers more sophisticated, and this dynamic might be interesting to study in a learning model. We studied the pricing problem of a monopolist seller of PDs; it would be useful to extend our framework to a competitive setting. Another fruitful avenue to research would be to compare and contrast the use of commitment devices that have received some attention in the current literature (Brocas et al., 2004b) with the use of PDs that are the focus of our study. From a normative perspective, although commitment devices (that are intended to moderate vice consumption) are compelling tools to curb the excesses associated with the self-control failures, their practical efficacy remains elusive. The purchase of commitment devices requires sophistication on the part of consumers, and extensive research has demonstrated that consumers often systematically are overconfident and hence are likely to end with a sub-optimal level of commitment (Heidhues & Kőszegi, 2009). This requirement is perhaps the reason that although researchers have paid more attention to commitment devices, policymakers seem to focus more heavily upon the PDs that are the focus of this study. Some recent work has also wondered why commitment devices that make eminent theoretical sense remain “missing” in practice (Laibson, 2015) and have indicated that their costs often overwhelm their perceived benefits. These costs could be direct (like price) but are often indirect, such as the loss of flexibility. For example, in a financial planning scenario, if consumers who want not to be tempted to misspend their money could put their savings into a perfectly illiquid savings account (that cannot be tapped at all before a desirable date- an example of a perfect commitment). But by doing so, they would lose all flexibility and hence would not be able to access these funds even in unforeseen emergencies. The absence of more flexible commitment devices in the field underscores the need to study alternative means of committing, and as such, our study is an attempt in that direction.

In our welfare analysis, we demonstrated that the compulsory provision of PDs could enhance the welfare of some segments while hurting others. As such, the impact of such subsidies also needs to consider the cost of provision for such subsidies. As a final note, we feel that an empirical examination of our findings through experimental and market data not only could serve as a validation test for our theory but also has potential to meaningfully amend and extend our model in more fruitful directions, leading both to better understanding of consumers and the potential for better public policy interventions.

Appendix A. Technical appendix

A.1. Proof of Proposition 1

Referring to Fig. 5 in the main text, notice that consumers in region R1 and R2 consume high and impose externality costs upon the society (absent a PD). Under a compulsory PD enforcement, they still consume high but no longer impose externality (provided they adopt PD). Hence, the total externality costs are as follows:

A. No uniform enforcement (R1 and R2)

$$E_{NE} = -e\frac{\Delta U - pI}{C_H}\left(1 - \frac{\Delta U - pI}{C_H}\right)$$

B. Uniform enforcement (R1, and R2)

$$E_{UE} = -e\left[(1-\tau)\frac{\Delta U - pI}{C_H}\left(1 - \frac{\Delta U - pI}{C_H}\right)\right],$$

where τ is the adoption rate of the preventative device. The difference in externality costs between A and B is:

$$\Delta E = E_{UE} - E_{NE} = -e\left[-\tau\frac{\Delta U - pI}{C_H}\left(1 - \frac{\Delta U - pI}{C_H}\right)\right] = e\tau\left[\frac{\Delta U - pI}{C_H}\left(1 - \frac{\Delta U - pI}{C_H}\right)\right] > 0$$

Thus, the externality-related costs are higher under no enforcement. In other words, the total externality-based criterion improves welfare under uniform enforcement.

Again, refer to Fig. 5 in the paper to consider the effects of uniform enforcement on the consumers. Notice that R1 and R2 consume high under both the regimes, but under compulsory PD, they do not put themselves under short-term danger (provided they adopt), and this improves their welfare. Those adopting from R4 and R5, end up consuming more, and this decreases their welfare.

Hence the utility difference between uniform enforcement and no uniform enforcement are derived as follows:
R1 and R2: $(\beta U_H - \beta C_H - T) - (\beta U_H - \beta pI - \beta C_H) = \beta pI > 0$
R4 and R5: $(\beta U_H - \beta C_H - T) - \beta U_L = \beta(\Delta U - C_H) < 0$

$$W_{12} = \int_0^{\frac{\Delta U - pI}{C_H}} \tau\beta pI \int_{\frac{\Delta U - pI}{C_H}}^{1} d\hat{\beta}\, d\beta = \tau pI\left(1 - \frac{\Delta U - pI}{C_H}\right)\frac{1}{2}\left(\frac{\Delta U - pI}{C_H}\right)^2$$

$$W_{45} = \int_{\frac{\Delta U - pI}{C_H}}^{\frac{\Delta U}{C_H}} \tau\beta(\Delta U - C_H) \int_{\beta}^{1} d\hat{\beta}\, d\beta = \int_{\frac{\Delta U - pI}{C_H}}^{\frac{\Delta U}{C_H}} \tau\beta(\Delta U - C_H)(1-\beta)\, d\beta$$

$$= \tau(\Delta U - C_H)\left\{\frac{1}{2}\left[\left(\frac{\Delta U}{C_H}\right)^2 - \left(\frac{\Delta U - pI}{C_H}\right)^2\right] - \frac{1}{3}\left[\left(\frac{\Delta U}{C_H}\right)^3 - \left(\frac{\Delta U - pI}{C_H}\right)^3\right]\right\}$$

The ***total utility-based welfare change*** as a result of the uniform enforcement of PD is given as:

$$\begin{aligned}
&W_{12} + W_{45} \\
&= \frac{1}{2}\tau pI\left(1 - \frac{\Delta U - pI}{C_H}\right)\left(\frac{\Delta U - pI}{C_H}\right)^2 + \frac{1}{2}\tau(\Delta U - C_H)\left(\frac{\Delta U}{C_H}\right)^2 - \frac{1}{2}\tau(\Delta U - C_H)\left(\frac{\Delta U - pI}{C_H}\right)^2 - \frac{1}{3}\tau(\Delta U - C_H)\left(\frac{\Delta U}{C_H}\right)^3 \\
&\qquad + \frac{1}{3}\tau(\Delta U - C_H)\left(\frac{\Delta U - pI}{C_H}\right)^3 \\
&= \frac{1}{2}\tau\left[\Delta U\left(\frac{\Delta U - pI}{C_H}\right)^2\left(1 - \frac{\Delta U - pI}{C_H}\right) - \frac{\Delta U - pI}{C_H}C_H\left(\frac{\Delta U - pI}{C_H}\right)^2\left(1 - \frac{\Delta U - pI}{C_H}\right)\right] \\
&\qquad + \frac{1}{2}\tau(\Delta U - C_H)\left(\frac{\Delta U}{C_H}\right)^2 - \frac{1}{2}\tau(\Delta U - C_H)\left(\frac{\Delta U - pI}{C_H}\right)^2 - \frac{1}{3}\tau(\Delta U - C_H)\left(\frac{\Delta U}{C_H}\right)^3 + \frac{1}{3}\tau(\Delta U - C_H)\left(\frac{\Delta U - pI}{C_H}\right)^3 \\
&= \frac{1}{2}\tau\left[\Delta U\left(\frac{\Delta U - pI}{C_H}\right)^2 - \Delta U\left(\frac{\Delta U - pI}{C_H}\right)^3 - C_H\left(\frac{\Delta U - pI}{C_H}\right)^3 + C_H\left(\frac{\Delta U - pI}{C_H}\right)^4\right] \\
&\qquad - \frac{1}{2}\tau(\Delta U - C_H)\left(\frac{\Delta U - pI}{C_H}\right)^2 + \frac{1}{3}\tau(\Delta U - C_H)\left(\frac{\Delta U - pI}{C_H}\right)^3 + \frac{1}{2}\tau(\Delta U - C_H)\left(\frac{\Delta U}{C_H}\right)^2 - \frac{1}{3}\tau(\Delta U - C_H)\left(\frac{\Delta U}{C_H}\right)^3 \\
&= \tau\left[\frac{1}{2}C_H\left(\frac{\Delta U - pI}{C_H}\right)^4 - \left(\frac{1}{2}\Delta U + \frac{1}{2}C_H - \frac{1}{3}\Delta U + \frac{1}{3}C_H\right)\left(\frac{\Delta U - pI}{C_H}\right)^3 + \left(\frac{1}{2}\Delta U - \frac{1}{2}\Delta U + \frac{1}{2}C_H\right)\left(\frac{\Delta U - pI}{C_H}\right)^2\right. \\
&\qquad \left. + \frac{1}{2}(\Delta U - C_H)\left(\frac{\Delta U}{C_H}\right)^2 - \frac{1}{3}(\Delta U - C_H)\left(\frac{\Delta U}{C_H}\right)^3\right] \\
&= \tau\left[\frac{1}{2}C_H\left(\frac{\Delta U - pI}{C_H}\right)^4 - \left(\frac{1}{6}\Delta U + \frac{5}{6}C_H\right)\left(\frac{\Delta U - pI}{C_H}\right)^3 + \left(\frac{1}{2}C_H\right)\left(\frac{\Delta U - pI}{C_H}\right)^2 + (\Delta U - C_H)\left(\frac{\Delta U}{C_H}\right)^2\left(\frac{1}{2} - \frac{1}{3}\frac{\Delta U}{C_H}\right)\right]
\end{aligned}$$

The ***total welfare change*** as a consequence of PD enforcement (including both the internalities and externalities) can be written as:

$$\Delta W = \tau\left[\frac{1}{2}C_H\left(\frac{\Delta U - pI}{C_H}\right)^4 - \left(\frac{1}{6}\Delta U + \frac{5}{6}C_H\right)\left(\frac{\Delta U - pI}{C_H}\right)^3 + \left(\frac{1}{2}C_H\right)\left(\frac{\Delta U - pI}{C_H}\right)^2 + (\Delta U - C_H)\left(\frac{\Delta U}{C_H}\right)^2\left(\frac{1}{2} - \frac{1}{3}\frac{\Delta U}{C_H}\right)\right]$$
$$+ e\tau\left[\frac{\Delta U - pI}{C_H}\left(1 - \frac{\Delta U - pI}{C_H}\right)\right]$$
$$= \tau\left[\frac{1}{2}C_H\left(\frac{\Delta U - pI}{C_H}\right)^4 - \left(\frac{1}{6}\Delta U + \frac{5}{6}C_H\right)\left(\frac{\Delta U - pI}{C_H}\right)^3 + \left(\frac{1}{2}C_H\right)\left(\frac{\Delta U - pI}{C_H}\right)^2 + (\Delta U - C_H)\left(\frac{\Delta U}{C_H}\right)^2\left(\frac{1}{2} - \frac{1}{3}\frac{\Delta U}{C_H}\right)\right]$$
$$+ e\tau\left[\frac{\Delta U - pI}{C_H} - \left(\frac{\Delta U - pI}{C_H}\right)^2\right]$$
$$= \tau\left[\frac{1}{2}C_H\left(\frac{\Delta U - pI}{C_H}\right)^4 - \left(\frac{1}{6}\Delta U + \frac{5}{6}C_H\right)\left(\frac{\Delta U - pI}{C_H}\right)^3 + \left(\frac{1}{2}C_H - e\right)\left(\frac{\Delta U - pI}{C_H}\right)^2 + e\frac{\Delta U - pI}{C_H} + (\Delta U - C_H)\left(\frac{\Delta U}{C_H}\right)^2\left(\frac{1}{2} - \frac{1}{3}\frac{\Delta U}{C_H}\right)\right]$$
$$= \tau\left\{\left[\frac{1}{2}C_H X^4 - \left(\frac{1}{6}\Delta U + \frac{5}{6}C_H\right)X^3 + \left(\frac{1}{2}C_H - e\right)X^2 + eX + M\right]\right\} \quad \text{(T1)}$$

where $M = (\Delta U - C_H)(\frac{\Delta U}{C_H})^2(\frac{1}{2} - \frac{1}{3}\frac{\Delta U}{C_H})$, and $X = \frac{\Delta U - pI}{C_H}$.

Note that, $M = (\Delta U - C_H)(\frac{\Delta U}{C_H})^2(\frac{1}{2} - \frac{1}{3}\frac{\Delta U}{C_H}) < 0$ since $\frac{\Delta U}{C_H} < 1$. Therefore, $\Delta W < 0$ at $X = 0$ when $X = \frac{\Delta U}{C_H}$,

$$\Delta W = \tau\left\{\left[\frac{1}{2}C_H X^4 - \left(\frac{1}{6}\Delta U + \frac{5}{6}C_H\right)X^3 + \left(\frac{1}{2}C_H - e\right)X^2 + eX + M\right]\right\}$$
$$= \tau\left\{\left[\frac{1}{2}\frac{\Delta U^4}{C_H^3} - \left(\frac{1}{6}\Delta U + \frac{5}{6}C_H\right)\left(\frac{\Delta U}{C_H}\right)^3 + \left(\frac{1}{2}C_H - e\right)\left(\frac{\Delta U}{C_H}\right)^2 + e\frac{\Delta U}{C_H} + M\right]\right\}$$
$$= \tau\left\{\left[\left(\frac{1}{3}\Delta U - \frac{5}{6}C_H\right)\left(\frac{\Delta U}{C_H}\right)^3 + \left(\frac{5}{6}\Delta U - e - \frac{1}{3}\frac{\Delta U^2}{C_H}\right)\left(\frac{\Delta U}{C_H}\right)^2 + e\frac{\Delta U}{C_H}\right]\right\} = \tau\left\{\left[e\frac{\Delta U}{C_H} - e\left(\frac{\Delta U}{C_H}\right)^2\right]\right\} > 0.$$

Since ΔW is continuous in X and has different signs at the upper and lower bounds of X, then there exists $0 < X^* < \frac{\Delta U}{C_H}$ such that $\Delta W = 0$ at $X = X^*$. Then we must have $\Delta W < 0$ when $0 < X < X^*$ and $\Delta W > 0$ when $\frac{\Delta U}{C_H} > X > X^*$.

Also note:

$$\frac{d(\Delta W)}{dX} = \tau\left\{\left[2C_H X^3 - \left(\frac{1}{2}\Delta U + \frac{5}{2}C_H\right)X^2 + (C_H - 2e)X + e\right]\right\} \quad \text{(T2)}$$

$$\frac{d^2(\Delta W)}{dX^2} = \tau\left\{\left[6C_H X^2 - (\Delta U + 5C_H)X + (C_H - 2e)\right]\right\}$$

$\Delta = (\Delta U + 5C_H)^2 - 24C_H(C_H - 2e) = C_H^2 + (\Delta U)^2 + 10\Delta U C_H + 48C_H e > 0$. (This is the radical from the quadratic root formula as applied to the second derivative expression).

The positive sign of the above expression guarantees that the second derivative of ΔW w.r.t. X has two positive real roots and can be written as: $\frac{d^2(\Delta W)}{dX^2} = \tau(X - X_1)(X - X_2)$.

Note that to prove Proposition 1 we need only the welfare change ΔW less than zero when X is less than X^* and ΔW greater than zero when X is greater than X^*.

Recall that $\Delta W < 0$ at $X = 0$ and $\Delta W < 0$ at $X = \frac{\Delta U}{C_H}$. Also, from Eq. (T1), we have $\Delta W \to +\infty$ when $X \to +\infty$ as well as when $X \to -\infty$. Then ΔW can have only the following two feasible shapes:

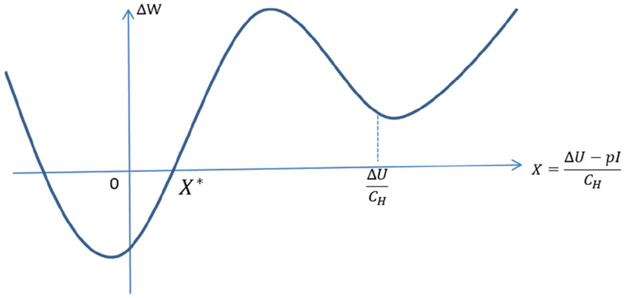

Fig. A1.

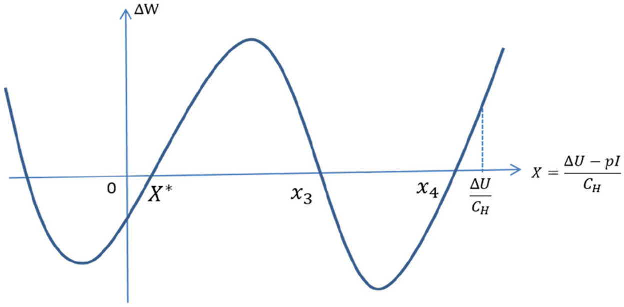

Fig. A2.

Fig. A1 is consistent with our results: $\Delta W < 0$ when $0 < X < X^*$ and $\Delta W > 0$ when $\frac{\Delta U}{C_H} > X > X^*$ regardless of the location of $\frac{\Delta U}{C_H}$.

In Fig. A2, since $\Delta W > 0$ when $X = \frac{\Delta U}{C_H}$, then, we have either (b1) $X^* < \frac{\Delta U}{C_H} < x_3$ or (b2) $\frac{\Delta U}{C_H} > x_4$.

Case b1. $X^* < \frac{\Delta U}{C_H} < x_3$ is consistent with P1 that $\Delta W < 0$ when $0 < X < X^*$ and $\Delta W > 0$ when $\frac{\Delta U}{C_H} > X > X^*$.

Case b2. When $\frac{\Delta U}{C_H} > x_4$, it contradicts P1 because $\Delta W < 0$ when $\frac{\Delta U}{C_H} > X > x_3 > X^*$.

We prove that such a shape is impossible by contradiction. Suppose ΔW is of this shape, then we have the relationships of X_1 and X_2, roots of $\frac{d^2(\Delta W)}{dX^2} = 0$ as follows:

$$0 < X_1 < \frac{\Delta U}{C_H},\ 0 < X_2 < \frac{\Delta U}{C_H},\ \text{and } X_1 + X_2 = \frac{5}{6} + \frac{1}{6}\frac{\Delta U}{C_H} < 2\frac{\Delta U}{C_H},$$

Therefore, $\frac{\Delta U}{C_H} > \frac{5}{11}$.

Since $X_2 = \frac{\Delta U + 5C_H + \sqrt{C_H^2 + (\Delta U)^2 + 10\Delta U C_H + 48 C_H e}}{12 C_H} < \frac{\Delta U}{C_H}$, we have $e < \frac{1}{2}\left(5\frac{\Delta U^2}{C_H} + C_H - 5\Delta U\right)$.

Since $e > 0$, then $\frac{1}{2}\left(5\frac{\Delta U^2}{C_H} + C_H - 5\Delta U\right) > 0$, thus we have

$$\frac{\Delta U}{C_H} > \frac{1}{2} + \frac{\sqrt{5}}{10} \tag{T3}$$

Since *we assume* ΔW has the shape as shown in Fig. A2, then we have $\frac{d(\Delta W)}{dX} > 0$ at $X = \frac{\Delta U}{C_H}$. Thus from Eq. (T2), we have $2C_H X^3 - \left(\frac{1}{2}\Delta U + \frac{5}{2}C_H\right)X^2 + (C_H - 2e)X + e > 0$ at $X = \frac{\Delta U}{C_H}$. After plugging in $X = \frac{\Delta U}{C_H}$, we have

$$\frac{\Delta U}{2}\left(3\frac{\Delta U^2}{{C_H}^2} + 2 - 5\frac{\Delta U}{C_H}\right) > \left(2\frac{\Delta U}{C_H} - 1\right)e$$

Then from Eq. (T3), we have $\frac{\Delta U}{2}\left(3\frac{\Delta U^2}{{C_H}^2} + 2 - 5\frac{\Delta U}{C_H}\right) > \left(2\frac{\Delta U}{C_H} - 1\right)e > 0$, thus we have $0 < \frac{\Delta U}{C_H} < \frac{2}{3}$, which contradicts Eq. (T3). Therefore, ΔW cannot be of the shape in Fig. A2.

A.2. Proof of Proposition 2

We can calculate the change in welfare as a result of a change in the perception of probability as follows:

$$\Delta S = S_2 - S_1 = -\frac{\Delta pI}{C_H}\left(\frac{\Delta U - pI - \Delta pI}{C_H} - 1 + \frac{\Delta U - pI}{C_H} - \frac{\Delta pI}{2C_H}\right) = \frac{\Delta pI}{C_H}\left(\frac{3\Delta pI}{2C_H} - 2\frac{\Delta U - pI}{C_H} + 1\right) = \frac{1}{C_H^2}\left[\frac{3\Delta pI^2}{2} - (2(\Delta U - pI) - C_H)\Delta pI\right] \tag{T4}$$

$$\frac{d\Delta S}{d\Delta pI} = \frac{1}{C_H^2}[3\Delta pI - 2(\Delta U - pI) + C_H] \text{ and}, \frac{d^2\Delta S}{d(\Delta pI)^2} = \frac{3}{C_H^2} > 0 \tag{T5}$$

Therefore, ΔS is convex in Δp. By equating the first derivative in Eq. (T5) to zero, we obtain: $\Delta p^* I = \frac{2\,(\Delta U - pI) - C_H}{3}$. There are two cases to be considered:

Case A. When $2\,(\Delta U - pI) - C_H > 0$, $\Delta p^* I = \frac{2\,(\Delta U - pI) - C_H}{3} > 0$. This case in shown in Fig. A3. Consumers worse off exceed those better off (i.e., $\Delta S < 0$) when $\Delta pI < \frac{2[2\,(\Delta U - pI) - C_H]}{3}$ (from Eq. (T4)) and this difference increases with Δp when $\Delta p < \Delta p^*$ and decreases with Δp when $\Delta p > \Delta p^*$. Consumers better off exceed those worse off (i.e., $\Delta S < 0$) when $\Delta pI > \frac{2[2\,(\Delta U - pI) - C_H]}{3}$ and this difference increases with Δp.

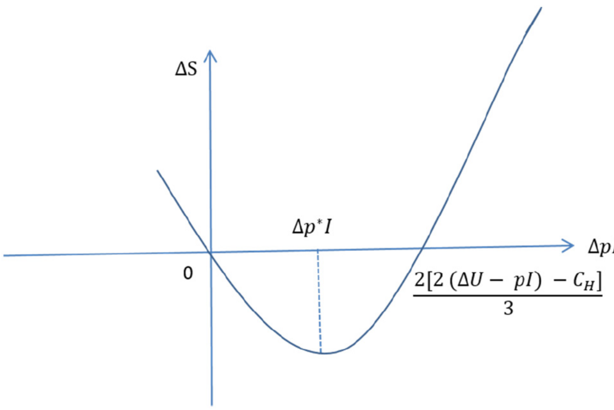

Fig. A3.

Therefore, policymakers will make more consumers worse off by slightly inflating the perceived probability $\hat{p}$. This inflation will be beneficial only when the inflation could exceed a threshold.

Case B. When $2\,(\Delta U - pI) - C_H < 0$, $\Delta p^* I = \frac{2\,(\Delta U - pI) - C_H}{3} < 0$, the outcome looks like the one in A4. Therefore, consumers better off always exceed those worse off (i.e., $\Delta S > 0$) for any $\Delta pI > 0$.

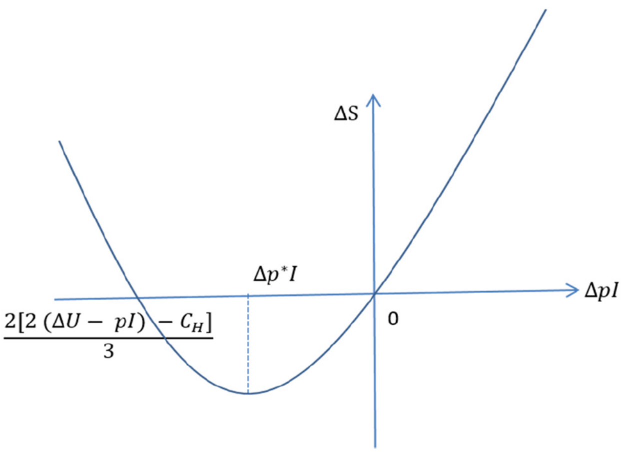

Fig. A4.

It is to be noted that assuming a relatively high value of I can guarantee $p + \Delta p$ is an appropriate probability measure.

A.3. Proof of Proposition 3

The expected revenue of the seller is given by:

$$\Pi = T \cdot\ Pr\left(\frac{T}{pI} < \hat{\beta} < \frac{\Delta U - pI}{C_H}\right) = T \cdot \frac{\frac{1}{2}\left(\frac{\Delta U - pI}{C_H} + \frac{T}{pI}\right)\left(\frac{\Delta U - pI}{C_H} - \frac{T}{pI}\right)}{\frac{1}{2}} = \left(\frac{\Delta U - pI}{C_H}\right)^2 T - \frac{T^3}{(pI)^2}$$

$$\frac{d\Pi}{dT} = \left(\frac{\Delta U - pI}{C_H}\right)^2 - \frac{3T^2}{(pI)^2} \text{ and, } \frac{d^2\Pi}{dT^2} = -\frac{6T}{(pI)^2}.$$

Notice $\frac{T}{pI} < \hat{\beta} < \frac{\Delta U - pI}{C_H}$ implies $T < \frac{\Delta U - pI}{C_H} pI$ where $\frac{\Delta U - pI}{C_H} < 1$. Since the second-order derivative is less than zero, then the first-order condition should yield the optimal solution. From the first-order condition, we have $T^* = \frac{\Delta U - pI}{\sqrt{3}C_H} pI$ and $\Pi^* = \frac{2}{3\sqrt{3}}\left(\frac{\Delta U - pI}{C_H}\right)^3 pI$.

$$\frac{dT^*}{dp} = \frac{I}{\sqrt{3}C_H}(-2pI + \Delta U) \tag{T6}$$

$$\frac{d^2T^*}{dp^2} = -\frac{2I^2}{\sqrt{3}C_H}$$

An examination of Eq. (T6) immediately yields the required result. Q.E.D.

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin*, *82*(4), 463.

Akerlof, G. A. (1991). Procrastination and obedience. *The American Economic Review*, *81*(2), 1–19.

Arcidiacono, P., Khwaja, A., & Ouyang, L. (2012). Habit persistence and teen sex: Could increased access to contraception have unintended consequences for teen pregnancies? *Journal of Business & Economic Statistics*, *30*(2), 312–325.

Ashraf, N., Karlan, D., & Yin, W. (2006). Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *The Quarterly Journal of Economics*, *121*(2), 635–672.

Baltimore Sun (2014). *Baltimore wants to hand out thousands more needles to drug users.* Accessed at: https://www.baltimoresun.com/politics/bs-xpm-2014-01-17-bs-md-srb-legislative-agenda-20140117-story.html.

Baumeister, R. F., & Vohs, K. D. (2003). Willpower, choice, and self-control. *Time and decision: Economic and psychological perspectives on intertemporal choice* (pp. 201–216).

BBC News (2006). Condom call for young 'on pull'. accessed at http://news.bbc.co.uk/go/pr/fr/-/2/hi/health/6132822.stm.

Bernheim, B. D., & Rangel, A. (2009). Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics. *The Quarterly Journal of Economics*, *124*(1), 51–104.

Brocas, I., Carrillo, J. D., & Dewatripont, M. (2004a). Commitment devices under self-control problems: An overview. *The Psychology of Economic Decisions*, *2*, 49–67.

Brocas, I., Carrillo, J. D., & Dewatripont, M. (2004b). Commitment devices under self-control problems: An overview. In I. Brocas, & J. D. Carrillo (Eds.), *Reasons and choices. Vol. 2.* (pp. 49–66).

Brown, A. L., Chua, Z. E., & Camerer, C. F. (2009). Learning and visceral temptation in dynamic saving experiments. *The Quarterly Journal of Economics*, *124*(1), 197–231.

Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). *Regulation for conservatives: Behavioral economics and the case for "asymmetric paternalism"*.

Campus Times (2013). Bring back the bar buses. (Published on Oct 13, 2013)

Carrillo, J. D., & Dewatripont, M. (2008). Promises, promises,…*. *The Economic Journal*, *118*(531), 1453–1473.

Castillo-Manzano, J. I., Castro-Nuño, M., & Pedregal, D. J. (2012). How many lives can bloody and shocking road safety advertising save? The case of Spain. *Transportation Research Part F: Traffic Psychology and Behaviour*, *15*(2), 174–187.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, *47*(2), 315–372.

DellaVigna, S., & Malmendier, U. (2004). Contract design and self-control: Theory and evidence. *The Quarterly Journal of Economics*, *119*(2), 353–402.

Dillard, J. P., Plotnick, C. A., Godbold, L. C., Freimuth, V. S., & Edgar, T. (1996). The multiple affective outcomes of AIDS PSAs fear appeals do more than scare people. *Communication Research*, *23*(1), 44–72.

Elster, J. (1979). *Ulysses and the sirens: Studies in rationality and irrationality*.

Gabaix, X., & Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *The Quarterly Journal of Economics*, *121*(2), 505–540.

Gilpatric, S. M. (2009). Slippage in rebate programs and present-biased preferences. *Marketing Science*, *28*(2), 229–238.

Gruber, J., & Kőszegi, B. (2004). Tax incidence when individuals are time-inconsistent: The case of cigarette excise taxes. *Journal of Public Economics*, *88*(9–10), 1959–1987.

Gul, F., & Pesendorfer, W. (2004). Self-control and the theory of consumption. *Econometrica*, *72*(1), 119–158.

Heidhues, P., & Kőszegi, B. (2009). Futile attempts at self-control. *Journal of the European Economic Association*, *7*(2–3), 423–434.

Heidhues, P., & Kőszegi, B. (2010). Exploiting naivete about self-control in the credit market. *The American Economic Review*, 2279–2303.

Jain, S. (2009). Self-control and optimal goals: A theoretical analysis. *Marketing Science*, *28*(6), 1027–1045.

Jain, S. (2012). Marketing of vice goods: A strategic analysis of the package size decision. *Marketing Science*, *31*(1), 36–51 2012.
Jain, S., & Li, K. J. (2018). Pricing and product design for vice goods: A strategic analysis. *Marketing Science*, *37*(4), 592–610.
Kaur, S., Kremer, M., & Mullainathan, S. (2015). Self-control at work. *Journal of Political Economy*, *123*(6), 1227–1277.
Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, *112*(2), 443–478.
Laibson, D. (2015). Why don't present-biased agents make commitments? *American Economic Review*, *105*(5), 267–272.
Las Vegas Sun (2003). Council holds off on designated drivers. Accessed at http://www.lasvegassun.com/news/2003/jun/19/council-holds-off-on-designated-drivers/.
Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, *65*(3), 272–292.
Meier, S., & Sprenger, C. D. (2013). Discounting financial literacy: Time preferences and participation in financial education programs. *Journal of Economic Behavior & Organization*, *95*, 159–174.
Milkman, K. L., Rogers, T., & Bazerman, M. H. (2009). Highbrow films gather dust: Time-inconsistent preferences and online DVD rentals. *Management Science*, *55*(6), 1047–1059.
Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., & Sears, M. R. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, *108*(7), 2693–2698.
Morgan, S. E., Palmgreen, P., Stephenson, M. T., Hoyle, R. H., & Lorch, E. P. (2003). Associations between message features and subjective evaluations of the sensation value of antidrug public service announcements. *Journal of Communication*, *53*(3), 512–526.
New York Times (2011). *U.S. releases graphic images to deter smokers.* Accessed at: https://www.nytimes.com/2011/06/22/health/policy/22smoke.html.
O'Donoghue, T., & Rabin, M. (1999). Doing it now or later. *American Economic Review*, *89*(1), 103–124.
O'Donoghue, T., & Rabin, M. (2000). The economics of immediate gratification. *Journal of Behavioral Decision Making*, *13*(2), 233–250.
O'Donoghue, T., & Rabin, M. (2006). Optimal sin taxes. *Journal of Public Economics*, *90*(10—11), 1825–1849.
Peltzman, S. (1975). The effects of automobile safety regulation. *The Journal of Political Economy*, 677–725.
Phelps, E. S., & Pollak, R. A. (1968). On second-best national saving and game-equilibrium growth. *The Review of Economic Studies*, *35*(2), 185–199.
Schaefer, R., Rao, R., & Mahajan, V. (2018). Marketing self-improvement program for self-signaling consumers. *Marketing Science*, *37*(6), 912–929.
Telegraph (2009). Government campaign says condoms make women sexier. Accessed at http://www.telegraph.co.uk/news/newstopics/howaboutthat/5020731/Government-campaign-says-condoms-make-women-sexier.html.
Telegraph (2015). *New drink-drive campaign launched.* Accessed at: https://www.telegraph.co.uk/motoring/road-safety/12027082/New-drink-drive-campaign-launched.html.
Thaler, R. H., & Shefrin, H. M. (1981). An economic theory of self-control. *The Journal of Political Economy*, 392–406.
Wang, Y., Lewis, M., & Singh, V. (2016). The unintended consequences of counter-marketing strategies: How particular anti-smoking measures may shift consumers to more dangerous cigarettes. *Marketing Science*, *35*(1), 55–72.
Wertenbroch, K. (1998). Consumption self-control by rationing purchase quantities of virtue and vice. *Marketing Science*, 317–337.