Harm Reduction for Addictive Consumption: When Does it Improve Health and When Does it Backfire?*

John Cawley† Cornell University

Davide Dragone‡ Università di Bologna

Abstract

Some harm reduction strategies encourage individuals to switch from a harmful addictive good to a less harmful addictive good; examples include e-cigarettes (substitutes for combustible cigarettes) and methadone and buprenorphine (substitutes for opioids). These have proven to be controversial. Advocates argue that people struggling with addiction benefit because they can switch to a less harmful substance, but opponents argue that this could encourage abstainers to begin using the harm-reduction method or even, eventually, the original addictive good. This paper builds on theories of addiction to model the introduction of a harm reduction method, and demonstrates the conditions under which each side is correct.

Keywords: addiction, dual-self, harm reduction, initiation, risk compensation, temptation
JEL codes: I12, I18, D11

*For helpful comments, we thank Catarina Goulao, Martin Forster, Don Kenkel, Catherine Maclean, Jesse Matheson, Michael Pesko, Luca Piccoli, Pedro Pita Barros, Heather Royer, Holger Strulik, Jenny Williams, and participants in the following seminars and conferences: the University of Bergamo seminar, the University of Bozen-Bolzano seminar, the University of Duisburg-Essen seminar, the University of Verona, the Toulouse School of Economics seminar, the Essen Health Conference, the Erasmus University of Rotterdam seminar, the EuHEA seminar, and the ASHEcon conference. We also thank the referees for their helpful feedback and suggestions.

†Cornell University, Brooks School of Public Policy; 2312 MVR Hall, Cornell University, Ithaca NY 14853, USA. e-mail: jhc38@cornell.edu

‡Corresponding author. Università di Bologna, Department of Economics, Piazza Scaravilli 2, 40126, Bologna, Italy; e-mail: davide.dragone@unibo.it

1 Introduction

Due to the substantial morbidity and mortality attributable to cigarette smoking, alcohol use disorder, and drug use disorder, nations worldwide have sought methods of reducing the health consequences of such addictive behaviors.[1] Governments have tried a zero-tolerance approach, which seeks to eliminate use of addictive drugs by relying heavily on law enforcement and less on treatment and rehabilitation. In contrast, the harm reduction approach accepts that some level of addictive consumption will always occur and focuses on reducing the health harms associated with it (Erickson, 1995; Single, 1995; Harm Reduction International, 2022; SAMHSA, 2023b). One example is Medications for Opioid Use Disorder (MOUD), which seeks to avoid the overdoses and other problems associated with fentanyl and heroin by providing people who use drugs (PWUD) with access to less addictive and less dangerous substances such as methadone and buprenorphine (National Harm Reduction Coalition, 2023).

Harm reduction methods such as MOUD have proven to be controversial, with advocates touting the potential health benefits, but opponents arguing that harm reduction methods that are themselves addictive could prevent users from quitting and may lead abstainers to begin using the harm reduction method, or even, eventually, the original addictive good.

The tradeoffs associated with harm reduction methods can be profitably studied by means of economic models. Specifically, in order to shed light on this debate, we propose a novel dynamic theoretical model to predict the consumption path of an addictive good, both before and after the introduction of a method of harm reduction. We then demonstrate the conditions under which the introduction of a method of harm reduction has the following consequences: it increases or decreases health harms; it leads previous users to quit the original addictive good; it leads previous abstainers to begin using the harm reduction method; or it leads previous abstainers to begin using the original addictive good. We demonstrate that the conditions for these different outcomes depend on three key factors: the enjoyableness of the harm reduction method; the addictiveness of the harm reduction method; and the substitutability of the harm reduction method with the original addictive good. To our knowledge, this is the first application of economic theory to harm reduction methods, and the first to demonstrate the conditions described above regarding the consequences of introducing harm reduction methods.

Common economic justifications for taxing or regulating any good (including harm reduction methods) include the neoclassical rationale of addressing negative externalities, and the behavioral economics rationale of addressing "internalities" that arise because of temptation and self-control costs. To allow for this latter possibility, we extend the benchmark model and

[1] The World Health Organization estimates that, worldwide, there are 8 million deaths annually from smoking, 3.3 million deaths annually from alcohol use disorder, and 500,000 deaths annually from drug overdose (WHO, 2022a,b,c).

consider a dual-self individual who makes farsighted decisions, but pays temptation costs to resist the myopic desire for immediate gratification (Gul and Pesendorfer, 2004; Loewenstein and O'Donoghue, 2004; Fudenberg and Levine, 2006; Gul and Pesendorfer, 2007). Our main results hold. Moreover, we show that the existence of temptation costs can lead to a consumer demand for policies that help them consume as they would in the benchmark case, in the absence of temptation costs.

This paper relates to several literatures. First, by deriving the conditions under which harm reduction leads to increased consumption by users or increased initiation by abstainers, we contribute to the economic literature on risk compensation in health behaviors (e.g. Cawley and Ruhm, 2012; Margolis et al., 2014; Dragone and Ziebarth, 2017; Simon et al., 2017; Toxvaerd, 2019; Frio and França, 2021; Doleac and Mukherjee, 2022). In general, this literature tends to find that even users of addictive substances respond to changes in risks and tradeoffs of consumption. Second, we contribute to the economic literature on the taxation and regulation of e-cigarettes (ENDS) and whether they are a substitute for combustible cigarettes (e.g. Friedman, 2015; Abouk and Adams, 2017; Marti et al., 2019; Pesko and Currie, 2019; Pesko et al., 2020; Allcott and Rafkin, 2022; Cotti et al., 2022; Pesko and Warman, 2022; Abouk et al., 2023). The findings of this literature generally suggest that cigarettes and e-cigarettes are to some extent substitutes, and that heavier regulation or taxation of e-cigarettes may lead to people consuming more combustible cigarettes. Third, we contribute to the literature regarding the regulation of methadone, buprenorphine, and naloxone (e.g. Abouk et al., 2019; Rees et al., 2019; Maclean et al., 2021; Barrette et al., 2021; Doleac and Mukherjee, 2022; Smart et al., 2023). Fourth, we contribute to the theoretical literature on addiction and self-control (Becker and Murphy, 1988; Gul and Pesendorfer, 2004; Loewenstein and O'Donoghue, 2004; Fudenberg and Levine, 2006; Gul and Pesendorfer, 2007; Dragone and Raggi, 2018, 2021).

2 Harm Reduction

The term harm reduction has been applied to a wide range of approaches, including syringe exchange programs, supervised injection facilities, legalized prostitution, condom distribution, Naloxone access laws, and Good Samaritan Laws (Stancliff et al., 2012; Rees et al., 2019; Doleac and Mukherjee, 2022; Packham, 2022; SAMHSA, 2023a; U.S. Department of Health and Human Services, 2023).

Importantly, for the purposes of this paper, we focus on the subset of harm reduction methods that involve the introduction of a substitute addictive good which is intended to reduce health harms. The mechanism of substitution may be that the harm reduction method binds to the same receptors in the brain as the original addictive good; in such a case the harm

reduction method is known as agonist therapy. Full agonists provide roughly the same euphoric effects as the original addictive good, whereas partial agonists are less euphoric (but still reduce feelings of withdrawal). Methadone (a full agonist) is to some extent a substitute for heroin, and buprenorphine (a partial agonist) is to some extent a substitute for opioid pain relievers. In both cases, the harm reduction method binds to similar opioid receptors in the brain as the original addictive good, leading to the release of similar neurotransmitters. In this sense, one can interpret the demand for both the original addictive good and the harm reduction method as a derived demand (Marshall, 1890; Lillard, 2020); i.e. derived from the demand for elevated levels of neurotransmitters associated with feelings of pleasure and reward, such as dopamine. Note that the harm reduction method may still be harmful to health, just not as harmful as the original product. (Obviously, if it was more harmful then it would not be a harm reduction method.)

There are numerous examples of such harm reduction methods. When the concern is the smoking of combustible cigarettes, harm reduction methods include electronic nicotine delivery systems or ENDS (commonly called e-cigarettes, the use of which is called vaping), and nicotine replacement therapy or NRT (which includes nicotine gum, patches, and lozenges). ENDS and NRT are harm reduction methods for combustible cigarettes because they are substitutes (they bind with the nicotine receptors and thus can reduce withdrawal from combustible cigarettes) and are believed to be less carcinogenic and toxic than cigarette smoke (although not likely completely safe). An additional harm reduction method to combat smoking that is popular in Norway and Sweden is snus, an oral tobacco product (Clarke et al., 2019).

When the concern is opioid addiction, relevant harm reduction methods include Opioid Agonist Therapy (OAT), which involves methadone or buprenorphine; these are a subset of Medications for Opioid Use Disorder (MOUD). When the concern is the smoking of combustible marijuana, harm reduction methods include edibles containing THC (a cannabinoid that provides a high), which allows the user to consume marijuana without inhaling toxic and carcinogenic smoke. Table 1 provides examples of harm reduction for cigarettes, heroin, opioid pain relievers, and marijuana, explaining why the harm reduction methods are substitutes for the original addictive substance, and how they may reduce harm.

Although we focus on a subset of harm reduction methods - those that are themselves somewhat addictive - in many ways these are the most interesting ones to examine, as they may be the most likely to involve unintended consequences and to be the most controversial. Our model, however, can be easily extended to account for harm reduction methods that are non addictive and non harmful.

Original Addictive Substance	**Harm Reduction Method**	**Why they are Substitutes**	**Why Harm (Potentially) Reduced**	**Possible Unintended Consequences**
Combustible Cigarettes	Nicotine replacement therapy or NRT (e.g. nicotine gum, patches, and lozenges); snus;	All bind with and activate nicotine receptors	NRT and snus do not involve smoking or inhaling anything, so are less carcinogenic	Current smokers may switch to e-cigarettes rather than quit altogether
	Electronic nicotine delivery systems or ENDS (aka e-cigarettes)		ENDS are more controversial; still involve inhaling chemicals but are believed to be less carcinogenic than regular smoking	Non-smokers may begin to vape; children a particular concern
	Chantix (varenicline)	Chantix is a partial agonist		Worst case: previous nonsmokers initiate smoking because of gateway effect of vaping
Heroin	Methadone (Opioid agonist therapy or OAT)	Both bind with and activate opioid receptors	Overdose is less likely on methadone than heroin, but still possible	Heroin addicts may switch to methadone rather than quit altogether
		Methadone activates receptors more slowly and less strongly so is less euphoric; reduces withdrawal but provides less high. Methadone is still considered a full agonist	Consumed orally rather than injected; less risk of HIV transmission	
Prescription opioid pain relievers	Buprenorphine (Opioid agonist therapy or OAT)	Both bind with and activate opioid receptors	Because buprenorphine is a partial agonist, overdose is less likely	Those addicted to opioid pain relievers may switch to buprenorphine rather than quit altogether
		Buprenorphine activates receptors more slowly and less strongly (it is a partial agonist) so is less euphoric; reduces withdrawal but provides less high. Buprenorphine is a partial agonist		Buprenorphine pills may be diverted for illicit use; may attract new users of opioids
Combustible marijuana	Edibles with THC	Both bind with and activate cannabinoid receptors	Consumed orally rather than inhaled, so are less carcinogenic	Current marijuana smokers may switch to edibles rather than quit altogether
			Concern about edibles being accidentally consumed, particularly by children	People who would not smoke marijuana may begin consuming edibles

Table 1: Examples of Harm Reduction Methods

Advocates of harm reduction[2] argue that it may help people quit, or at least reduce their consumption of, the original addictive good, and may reduce health harms. Opponents are concerned that harm reduction methods may decrease quitting of the original addictive good, because the very harmfulness of the original addictive good provides an incentive to quit. If, however, a substitute product is made available that is less harmful then users may switch to that substitute rather than quit altogether. Opponents are also concerned that harm reduction methods may increase addiction. Some people may be abstaining precisely because the original substance is harmful to health; introducing a product that is less harmful may encourage some of those who previously abstained to begin using the new harm reduction method (for example, some non-smokers may begin vaping.) Even worse, some of those previous abstainers who begin using the harm reduction method may eventually transition to the original, more harmful, addictive good.[3]

The arguments of both sides relate to risk compensation and moral hazard. In most countries, the majority of health care costs are paid by third-party payers, such as public health insurance programs and private (commercial) health insurance. Thus, consumer responses to the introduction of the harm reduction method that worsens health harms (such as increased use of the original addictive good, or use of the harm reduction method by previous abstainers) impose negative externalities through health insurance and represent a form of ex-ante moral hazard (Ehrlich and Becker, 1972; Einav and Finkelstein, 2018). This is akin to a situation in which improvements in car safety lead drivers to take greater risks, potentially increasing negative externalities (Peltzman, 1975). The final column of Table 1 provides examples of risk compensation and unintended consequences for each of the examples of harm reduction.

The ambivalence about harm reduction affects many aspects of regulation and policy. Physicians in the U.K., Canada, and Australia have been able to prescribe take-home methadone since the 1960s, but in the U.S. the substance is much more restricted - patients are generally required to travel to methadone clinics where staff must observe them ingest the medication (Singer and Hamilton, 2023). Until 2023, buprenorphine was more heavily regulated in the U.S. than the opioid pain relievers that contributed heavily to the fatal drug overdose epidemic (Powell et al., 2020; Maclean et al., 2021; Alpert et al., 2022; SAMHSA, 2023c). Before 2023, in order to prescribe buprenorphine, physicians in the U.S. had to undertake 8 hours of training and obtain a waiver from the U.S. Drug Enforcement Agency, and even that only enabled them to prescribe it to a limited number of patients (Waters, 2019). As a result, 40% of U.S.

[2]Organizations in favor of liberalizing access to harm reduction methods include Harm Reduction International, the National Harm Reduction Coalition and the Drug Policy Alliance.

[3]For a discussion of the controversy over harm reduction methods, see Campbell (2009) regarding the UN declaration of intent toward harm reduction policies, Satel (2019) on ENDS, and Vestal (2016) and Singer and Hamilton (2023) on methadone.

counties had no physicians who could prescribe buprenorphine, and another 24% of counties had insufficient prescribing capacity (Grimm, 2020). A policy that continues to restrict access is that numerous states' Medicaid programs require prior authorization before they will cover the cost of buprenorphine (Weber and Gupta, 2019).

Likewise, electronic nicotine delivery systems (ENDS) are sometimes more tightly regulated than combustible cigarettes. The World Health Organization reports that 32 nations (including Australia, Brazil, India, Japan, and Mexico) have banned e-cigarettes entirely (WHO, 2022d). In countries where ENDS are legal, they range from completely unregulated to regulated as pharmaceutical products (WHO, 2022d). In 2022, the U.S. Food & Drug Administration (FDA) issued a marketing denial order to Juul, which banned them from selling any of their ENDS products in the United States, despite the fact that cigarettes remain legal to sell (U.S. Food and Drug Administration, 2022).[4]

There is also debate about the optimal taxation of e-cigarettes. In the U.S., 21 states do not tax e-cigarettes at all. Among those that do tax e-cigarettes, the structure and amount of those taxes vary considerably. Among states that tax e-cigarettes on the basis of their wholesale price, the tax rates range from 8% in New Hampshire to 95% in Minnesota (IGEN, 2022).

3 A Model of Harm Reduction and Addictive Consumption

We extend the model of rational addiction developed by Becker and Murphy (1988) and the extensions into cycles of addictive consumption by Dockner and Feichtinger (1993). We first consider the conditions for a person to become a consumer of an addictive good. We then introduce a harm reduction method, and show how this affects health and the consumption of the original addictive good. Analytical solutions of the problem are provided in the Appendices. We start with a rational choice model, but in Section 8 we modify it to allow a dual-process model, in which observed choices are the result of a disagreement between a rational forward-looking self and a myopic impulsive self.

3.1 Addictive Consumption in the Absence of a Harm Reduction Method

Define c as the consumption of an addictive good.[5] Consumption of this good contributes to two stocks. The first one is an addictive stock A, a measure of the past consumption experiences

[4] An appeals court entered a temporary stay of the FDA's marketing denial order, and the FDA is conducting additional review, but the FDA has clarified that this stay does not constitute authorization for Juul to market or sell their products (U.S. Food and Drug Administration, 2022).

[5] The addictive good may be legal or illegal. For the sake of simplicity we discuss only its pecuniary price; for an illegal substance one might also want to consider other aspects of total cost, including the time costs of acquisition, the expected probability of being arrested and the utility loss of being arrested (Becker, 1968).

with the addictive good, which evolves over time according to

$$\dot{A}(t) = c(t) - \delta_A A(t) \tag{1}$$

where $\delta_A > 0$ is the depreciation rate of the addictive stock. The second stock is H, which describes the negative health consequences of addictive consumption, i.e. health harm. Stock H increases with both current and past consumption of the addictive good according to Allcott and Rafkin (2022)

$$\dot{H}(t) = c(t) + \omega A(t) - \delta_H H(t) \tag{2}$$

where $\omega > 0$ is the marginal contribution of addiction to health harms and $\delta_H > 0$ is the depreciation rate of health harms. The change in health $\dot{H}$ depends on the level of addictive stock A because there may be cumulative health consequences to long-term addiction, via mechanisms such as exposure to infected needles or inhaling cigarette smoke. For this reason, the change in health depends not just on *current* addictive consumption c but also the *history* of addictive consumption as measured by the addictive stock A.

Consider the following utility function

$$\mathcal{U}(c, q; A, H) = \left(u_c + u_{cA}A + \frac{u_{cc}}{2}c\right)c + \left(u_A + \frac{u_{AA}}{2}A\right)A + \left(u_H + \frac{u_{HH}}{2}H + u_{AH}A\right)H + q \tag{3}$$

where u_A, u_H and $u_{cc}, u_{AA}, u_{HH}, u_{AH}$ are negative, and variable q represents a numeraire composite good. This linear-quadratic specification, which can be considered to be a second-order approximation of a more general utility function, is common in the rational addiction literature because it allows for a closed-form solution of the optimal trajectory and the steady state (see, e.g., Becker and Murphy, 1988; Chaloupka, 1991; Gruber and Köszegi, 2001; Dragone and Raggi, 2018, 2021; Piccoli and Tiezzi, 2021).[6] We allow for the interaction between A and H in order to be flexible (about both the dynamics of consumption and the utility function), but the results do not depend on this term.

We call parameter $u_c > 0$ the enjoyableness of the addictive good. It corresponds to the positive component of the marginal utility of consumption confronting someone who has until then abstained from c – i.e. the marginal utility they would experience from their first use of c. A defining assumption of the rational addiction model is that c is addictive, which is formalized assuming that the larger the stock of A the larger the marginal utility of c, i.e. $u_{cA} > 0$. This is the nature of addiction: the more of an addictive good one has consumed in the recent past, the greater one's marginal utility of consumption (or, put another way, the greater the withdrawal

[6] The quasi-linear specification rules out income effects to better focus on substitution patterns (Dragone and Vanin, 2022), as we do in this paper.

one experiences from not consuming it). In the rational addiction literature, this is referred to as adjacent complementarity or reinforcement (see, e.g., Becker and Murphy, 1988; Becker et al., 1991). For consistency, we refer to $u_{cA} > 0$ as the degree of addictiveness of the good.

Given a discount rate ρ, the intertemporal rational addiction problem of the agent is

$$\max \int_0^\infty e^{-\rho t} \mathcal{U}\left(c(t), q(t), A(t), H(t)\right) \mathrm{d}t \tag{4}$$

subject to the law of motions (1) and (2), and the budget constraint

$$M\left(t\right) = q\left(t\right) + p_c c\left(t\right) \tag{5}$$

where M is income and p_c describes the monetary price, possibly including taxes, of the addictive good.[7]

To determine the optimal quantity of addictive consumption, we apply the Pontryagin's maximum principle and obtain the optimal trajectory of consumption, addiction and health harm toward the steady state. As shown in Appendix A.1, the steady state level of consumption of the addictive good is

$$c^{ss} = \alpha\left(u_c - p_c - \pi_c\right) \tag{6}$$

with $\alpha > 0$. The term $\pi_c > 0$ represents the decrease in future utility resulting from consumption that raises addiction and health harms. As shown in equation (38) in the Appendix, the higher marginal disutility from addiction and health harms, the higher the value of π_c. A higher π_c implies a lower optimal level of consumption for the addictive good, as it increases the future costs associated with current consumption. Additionally, π_c is higher when the agent is more patient, as represented by a lower discount rate ρ. Therefore, a more patient agent will consume less of the addictive good to mitigate the negative effects of addiction and health harms on future utility. According to eq. (6), the addictive good is consumed only if the benefits exceed the costs – i.e. if the instantaneous marginal utility of consumption from first-time use exceeds the sum of the monetary costs of purchasing the addictive good and the

[7] Considering an infinite time horizon, instead of a finite one, is appropriate when an agent is not planning to die at a predetermined date T. Given the exponential discounting function, the conditional probability of dying at age t, given that the agent has not died before age t (i.e. the hazard rate) is ρ in each period, the expected time of death is $1/\rho$, and the probability of living forever is zero (see, e.g., Becker and Murphy, 1988; Mas-Colell et al., 1995; Dragone and Raggi, 2018, 2021). This implies that people make plans as if they could live indefinitely, even though they will actually die at a finite time. Note that, by Samuelson (1965)'s turnpike theorem, the solution of the infinite-horizon problem provides an approximation for the solutions of problems with a predetermined time of death. Extending the model to allow for an endogenous (state-dependent) survival probability is meaningful and realistic, but does not qualitatively alter our results. To ease the exposition, here we abstract from considerations about the time cost of obtaining the good due to, e.g. the time spent on consumption, the need for medical prescriptions, or expected sanctions and the risk of accessing the black market (if consumption of the good is illegal).

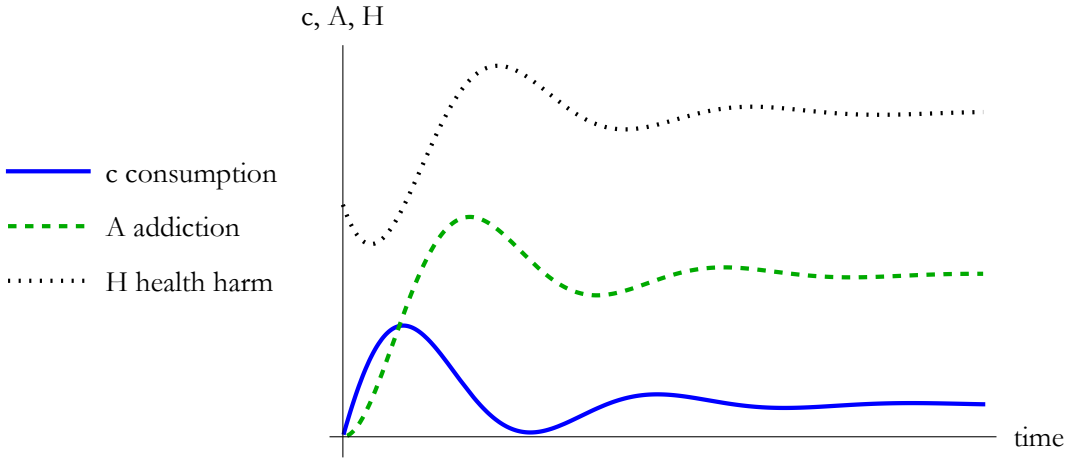

Figure 1: Initiation and addictive consumption when one addictive good is available. Illustration of possible time-paths for c, A and H. Parameters: $u_c = 6, u_A = u_H = -0.1, u_{cc} = u_{AA} = u_{HH} = -1, u_{cA} = 0.9, u_{HA} = 0, \delta_A = 0.1, \delta_H = 0.2, \omega = 0.5, \rho = 2.5, p_c = 0.1$.

future consequences of consumption.

In empirical applications, it may be useful to estimate how the optimal level of consumption of the original addictive good evolves over time, as a function of the state of addiction and health. This corresponds to estimating the parameters of the following policy function

$$\tilde{c}\,(A, H) = a_c c^{ss} + a_A A + a_H H \tag{7}$$

This function is the two-state analog of the policy function proposed by Dragone and Raggi (2021). Since it describes a stationary consumption trajectory, it is not subject to the empirical concerns raised by Auld and Grootendorst (2004) and Laporte et al. (2017) about the possible difficulties in estimating the traditional rational addiction model.

Only the steady state level of consumption c^{ss} of the policy function (7) depends on the market price (or tax) of the addictive good, hence $\partial\tilde{c}/\partial p_c = \partial(a_c c^{ss})/\partial p_c$, which is predicted to be negative. The estimation of this term allows measuring the response of addictive demand to exogenous shocks, such as tax changes or the introduction of bans (which can be considered as an infinitely high tax), or the price/tax elasticity of demand (Dragone and Raggi, 2018), conditional on the state of addiction and health harms.[8] Estimation of parameter a_H allows measuring the response to health shocks, or new information about own health condition (see, e.g., Smith et al., 2001; Cawley and Ruhm, 2012; Darden and Gilleskie, 2016; Arni et al., 2021).

Note that, despite being a linear equation in addiction and health harm, the time trajectory

[8] In the empirical literature on addictive consumption, the state of addiction is often proxied by lagged consumption of the addictive good (see, e.g., Chaloupka, 1991; Becker et al., 1994), or by the duration of the history of addictive consumption.

for consumption allows for oscillations as a possible consequence of the underlying dynamics of the state variables (see Figure 1 for an illustration). This implies that, even if consumption is zero in the long run, it does not preclude the possibility that the person had earlier experimented with the addictive good.

3.2 Addictive Consumption in the Presence of a Harm Reduction Method

We now introduce a harm reduction method v, which is an addictive and less-harmful substitute for the addictive good c. For example, if c describes cigarettes then v can describe vaping of e-cigarettes. The harm reduction method v adds to both the addictive stock A and the stock of health harms H

$$\dot{A}(t) = c(t) + \varepsilon_A v(t) - \delta_A A(t) \tag{8}$$

$$\dot{H}(t) = c(t) + \varepsilon_H v(t) + \omega A(t) - \delta_H H(t) \tag{9}$$

By definition, the harm reduction method is assumed to be less harmful to health than the original addictive good, hence $\varepsilon_H \in [0, 1)$.[9] In our model, the harm reduction can be more, equally, or less addictive than the original addictive good, which is useful because in practice they could be any of the three.[10] For example, partial agonists may be less addictive than the addictive goods they are meant to address. However, some harm reduction methods allow the consumer to choose the intensity; e.g. people generally consume less nicotine through e-cigarettes than they would through combustible cigarettes, but someone experienced with the technology could consume more (Prochaska et al., 2021).

When the harm reduction method is available, the agent's instantaneous utility function is

$$\mathcal{V}(c, v, q; A, H) = \mathcal{U}(c, q; A, H) + \left(u_v + u_{vc}c + u_{vA}A + \frac{u_{vv}}{2}v\right)v \tag{10}$$

where $\mathcal{U}(\cdot)$ is defined in (3) and $u_{vv} < 0$. Analogously to the benchmark case, the positive term u_v describes the enjoyableness of the harm reduction method, absent previous and current consumption, and $u_{vA} \geq 0$ describes the addictiveness of the harm reduction method (i.e.

[9] This is a reasonable assumption to make, but there have been times when, although the perceived harm was lower than the original addictive good, the actual harm was greater than the substance they were intended to replace. For example, heroin was originally marketed as a safe and non-addictive alternative to morphine, and OxyContin was originally marketed as a safer and less addictive alternative to older opioid pain relievers.

[10] There are harm reduction methods that decrease the marginal utility of the original addictive good, but are non-addictive. These include antagonists, which block rather than activate the receptors used by the original addictive good. Examples include the antagonists naltrexone (for opioids or alcohol), naloxone (for opioids), disulfiram (for alcohol) and acamprosate (for alcohol). Here, for clarity and focus we assume it is a full or partial agonist rather than an antagonist. However, as shown in Section 5.1, our model is flexible and allows the harm reduction method to be an antagonist.

the effect of past use on the marginal utility of current use). The term u_{vc} describes the substitutability (in preferences) between simultaneous consumption of the two addictive goods. It is negative if the harm reduction method is a substitute for the original addictive good (which is what we assume). In contrast, that term would be positive if the harm reduction method is a complement with the original addictive good.[11]

Overall, the use of the harm reduction method v can affect the consumption of the original addictive good through two channels. The first, direct, channel operates through individual preferences, in that v is enjoyable ($u_v > 0$), addictive ($u_{vA} > 0$) and affects the marginal utility of the original addictive good (through u_{vc}). The second, indirect, channel operates through the two stocks of addiction and health harms, as described by ε_A and ε_H in the law of motion of addiction and health harms.

The steady-state consumption of the harm reduction method can be described as follows

$$v_d^{ss} = \theta_v \left(u_v - p_v - \pi_v\right) + \gamma c^{ss} \tag{11}$$

where θ_v is positive, and $\pi_v > 0$ describes the decrease in future utility resulting from the harm reduction method raising the addictive stock A and health harms H (see Appendix A.3 for details).[12]

The logic of equation (11) is that people will consume the harm reduction method only if the benefits exceed the costs. In this case the benefits include the enjoyableness of the harm reduction method (u_v), while the costs include the monetary costs of purchasing the harm reduction method (p_v) and the future harmful consequences of consuming it (π_v). Moreover, due to the interdependence between the two addictive goods, use of the harm reduction method also depends on the consumption of the original addictive good.

Specifically, the term γ, which is multiplied by the steady-state level of consumption of the original addictive good (c^{ss}), is higher the greater u_{vA}, u_{cA} and u_{vc}, i.e. the greater the degree of complementarity between the goods and the addictive stock (see eq. 68). This occurs because both goods contribute to the accumulation of the addictive stock A, so the consumption of one good reinforces consumption of the other. On the contrary, the higher the substitutability between the original addictive good and the harm reduction method (i.e. low values of u_{vc}), the lower the mutually reinforcing effect of consuming c and v, and γ.

[11] Note that, by referring to substitutability in preferences, we consider a property of the utility function, as described by the cross-derivative u_{vc}. We are not referring to the definition of gross complements and substitutes, used to describe how the demand for one good responds to changes in the price of another good.

[12] Subscript d is a mnemonic of dual consumption (i.e., after the introduction of the harm reduction method). Recall that the benchmark steady-state values have no subscripts.

The new steady-state levels of consumption of addictive good and health harm are

$$c_d^{ss} = c^{ss} + \theta_c \left(u_{vA} - r_H\left(u_{vc}\right)\right) v_d^{ss} \tag{12}$$

$$H_d^{ss} = H^{ss} + \theta_H \left(u_{vA} - r_L\left(u_{vc}\right)\right) v_d^{ss} \tag{13}$$

where $\theta_c, \theta_H > 0$, and $r_L\left(u_{vc}\right) < r_H\left(u_{vc}\right)$ are threshold levels that depend on the degree of substitutability between v and c (see equations 69 and 70).

Equations (12) and (13) will be used in the next Section to illustrate the conditions under which the introduction of a harm reduction method increases or decreases the consumption of the original addictive good and the magnitude of health harms.

4 Consequences of the introduction of a harm reduction method

Just because a harm reduction method becomes available does not mean that people will use it. The model implies that people will use the harm reduction method in the long run when

$$u_v > p_v + \pi_v - \frac{\gamma}{\theta_v} c^{ss} \tag{14}$$

Based on the previous discussion about the determinants of γ, the following holds:

Remark 1 *The use of the harm reduction method is more likely the greater its enjoyableness and the lower its full price (which includes monetary price and future health harms). If c and v are sufficiently addictive and complements, then previous consumers of the original addictive good are more likely than previous abstainers to eventually use the harm reduction method.*

In what follows, we focus on scenarios in which the individual decides to use the harm reduction method (i.e. condition 14 is satisfied). We show that the consequences of using the harm reduction method depend on three key factors: the enjoyableness of the harm reduction method (u_v), the addictiveness of the harm reduction method (u_{vA}), and the substitutability of the harm reduction method for the original addictive good (u_{vc}). For later reference, consider the following terminology:[13]

Definition 1 *A harm reduction method is defined as*

- *Mildly addictive if $u_{vA} < r_L\left(u_{vc}\right)$;*
- *Moderately addictive if $u_{vA} \in \left(r_L\left(u_{vc}\right), r_H\left(u_{vc}\right)\right)$;*

[13] The precise expressions for r_L and r_H are reported in equations (77) and (78) in Appendix A.3.

- *Highly addictive if* $u_{vA} > r_H(u_{vc})$.

We will discuss the implications for two types of individuals. First, we consider the impacts on consumers of the original addictive good. This category of consumers (e.g. smokers or heroin users) is a main target for policies that aim to reduce health harms and addiction. Second, we consider the impacts on individuals who were abstaining from the addictive good before the harm reduction method was introduced. This group is of interest because their use of the harm reduction method does not entail any benefits in terms of reduced use of the original addictive good, and because of concerns that the harm reduction method could turn out to be a gateway, leading them to consume the original addictive good from which they previously abstained.

4.1 Effects on the consumption of the original addictive good: gateway effects, substitution effects, and quitting

To assess how the use of the harm reduction method affects consumption of the original addictive good, consider expression (12), which can be conveniently rewritten as

$$c_d^{ss} - c^{ss} = \theta_c \left(u_{vA} - r_H\left(u_{vc}\right)\right) v_d^{ss} \tag{15}$$

Consistent with intuition, equation (15) shows that, if the individual is not using the harm reduction method in the long run, the steady-state consumption of the original addictive good is unaffected and thus addiction and health remain unchanged.

Moreover, since θ_c is a positive parameter the following result holds:

Proposition 1 (Consumption of the original addictive good) *Conditional on using the harm reduction method, consumption of the original addictive good c increases in the long run if the harm reduction method v is highly addictive* ($u_{vA} > r_H$)*, and it decreases otherwise.*

Table 2 organizes the results for an individual already consuming the original addictive good prior to the introduction of a harm reduction method. The first row concerns the case where the individual uses the harm reduction method, and the second row concerns the case where the individual does not use it, after it has been introduced. Columns 1, 2, and 3 describe the cases where the addictiveness of the harm reduction method is mild, moderate, and high.

When the harm reduction method is highly addictive, steady-state consumption of the addictive good increases. Use of the harm reduction method increases the addictive stock A, which increases the marginal utility not just of the harm reduction method but also of the original addictive good. Incentivized by the higher marginal utility of consumption, the individual increases their consumption of the original addictive good. Hence, the harm reduction method

Harm reduction method used?	Addictiveness of harm reduction method: Mild	Moderate	High
Yes	**Harm reduction** $c^{ss} \downarrow$, $H^{ss} \downarrow$	**Substitution** $c^{ss} \downarrow$, $H^{ss} \uparrow$	**Harm reduction backfires** $c^{ss} \uparrow$, $H^{ss} \uparrow$
No	**Harm reduction method irrelevant** $v^{ss} = 0$; c^{ss} and H^{ss} remain unchanged		

Table 2: Long-run changes in consumption and health harm for an individual **already consuming** the addictive good prior to the introduction of a harm reduction method ($c^{ss} > 0$). If the harm reduction method v is enjoyable enough, harm reduction results when it is mildly addictive ($u_{vA} < r_L$), and substitution results when v is moderately addictive ($u_{vA} \in (r_L, r_H)$), and harm reduction backfires – the worst case scenario from the public health perspective – when v is highly addictive ($u_{vA} > r_H$).

backfires: use of v induces increased consumption of the original addictive good c. Overall, dual consumption and higher health harm result, as illustrated in the right-most panel of Figure 2.[14]

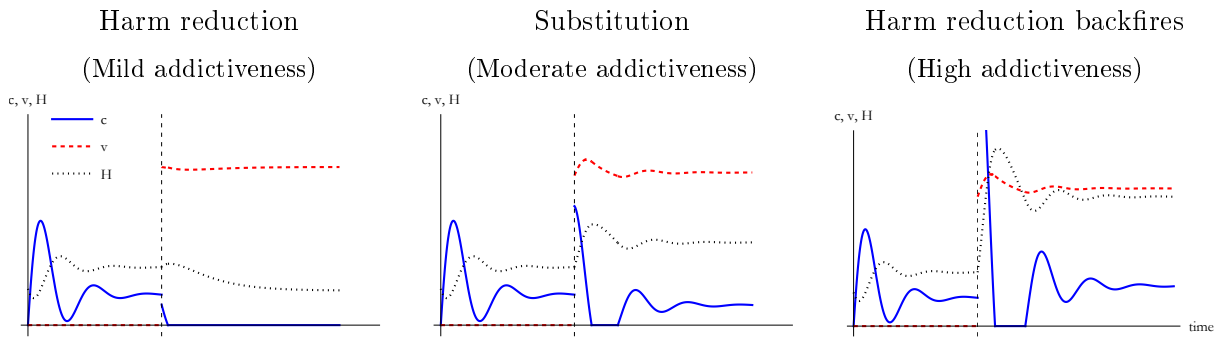

Figure 2: Illustration of some possible trajectories of consumption and health harm for an individual already consuming the addictive good c before the harm reduction method v is introduced. The vertical dashed line denotes when the method becomes available. Left panel: the consumer quits the original addictive good and health harm is reduced. Center panel: the harm reduction method substitutes for the original addictive good, health harm increases. Right panel: addictive consumption and health harm increase. In all panels, condition (14) is satisfied and the individual uses the harm reduction method. Parameters as for Figure 1, with $u_{vc} = -0.1, p_v = 1, \varepsilon_A = 0.1, \varepsilon_H = 0.3$. For the left panel: $u_v = 17, u_{vv} = -4, u_{vA} = 0.3$; for the central panel: $u_v = 80, u_{vv} = -23, u_{vA} = 2$; for the right panel: $u_v = 570, u_{vv} = -160, u_{vA} = 5.5$.

Remark 2 *When the harm reduction method is highly addictive, the availability and use of the harm reduction method backfires, because it induces higher consumption of the original addictive good.*

A noteworthy outcome related to Remark 2 arises when the introduction of the harm reduc-

[14] When the harm reduction method is not highly addictive, its use effectively replaces consumption of the original addictive good; they behave like substitutes.

tion method leads previous abstainers to initiate consumption of the original addictive good (see Table 3). In this case, the harm reduction method has become a gateway drug. It is the consequence of the harm reduction method being enjoyable and highly addictive which produces dual addictive consumption. This outcome is more likely, the lower the threshold r_H defining the harm reduction method as highly addictive.

Harm reduction method used?	Addictiveness of harm reduction method: Mild or Moderate	Addictiveness of harm reduction method: High
Yes	**Initiate only the harm reduction method** $c^{ss} = 0$, $H^{ss} \uparrow$	**Gateway effect: Initiate both the harm harm reduction method and addictive good** $c^{ss} \uparrow$, $H^{ss} \uparrow$
No	**Harm reduction method is irrelevant** $v^{ss} = c^{ss} = H^{ss} = 0$	

Table 3: Long-run changes in consumption and health harm for an individual **previously abstaining** from the addictive good c prior to the introduction of a harm reduction method. If the harm reduction method is used, health harm always increases. Gateway effects and initiation with both addictive goods result if v is highly addictive ($u_{vA} > r_H$): the harm reduction policy backfires.

Consider now the possibility that consumption of the original addictive good decreases. Based on the fact that $\partial r_H / \partial u_{vc} < 0$ (see equation 72), the following holds:

Remark 3 *Conditional on using the harm reduction method, the long-run consumption of the original addictive good is more likely to decrease:*

- *The lower the addictiveness of the harm reduction method;*
- *The greater the substitutability between the harm reduction method and the original addictive good.*

The two left-most panels of Figure 2 illustrate this case when v is only mildly or moderately addictive, so that consumption of the original addictive good (solid line) tends to decrease over time. Importantly, in such cases the harm reduction method can lead the individual to not only reduce consumption of the original addictive good but quit it altogether (i.e. $c^{ss} > c_d^{ss} = 0$).

Remark 4 (Quitting the original addictive good) *For an individual previously using the original addictive good, quitting is more likely*

a. *The lower the consumption of the original addictive good c^{ss};*

b. *The lower the addictiveness of the harm reduction method u_{vA};*

c. *The greater the use of the harm reduction method* v_d^{ss}.

Condition (a) states an intuitive condition: it is harder for a heavy smoker than for a light smoker to quit. The logic of condition (b) is that, all else equal, a less addictive harm reduction method promotes quitting because it contributes less to the addictive stock and thus does less (via adjacent complementarity) to increase the marginal utility of the original addictive good. Condition (c) is of particular interest, because there may be concern about the health harms due to high consumption of the harm reduction method, but the trade-off is that greater use of the harm reduction method increases the likelihood that previous users will quit the addictive good.

Figure 2 illustrates how an individual's adjustment to their consumption of the original addictive good varies depending on the addictiveness of the harm reduction method. In the case of mild addictiveness, the harm reduction method contributes minimally to the stock of addiction, reducing the incentive to consume the original addictive good. However, when the addictiveness of the harm reduction method is moderate, its contribution to the addiction stock is more substantial. This initially leads to higher consumption of the original addictive good. As health harm increases due to the initial period of dual consumption, the individual subsequently reduces their intake of the original addictive good, temporarily quitting. Once the health harm decreases sufficiently, consuming the original addictive good becomes more attractive, leading to a relapse. This relapse, in turn, increases the addiction stock, resulting in dual use, albeit with lower consumption of the original addictive good. In the case of high addictiveness of the harm reduction method, a similar dynamic emerges, but the relapse to consuming the original addictive good occurs at a higher level of health harm (and addiction) due to the large contribution of the harm reduction method to the addiction stock.

Another important insight from Figure 2 is that the introduction of a harm reduction method can cause a short-run increase in the consumption of the original addictive good, even if the steady state consumption of the original addictive good will eventually be zero. Importantly, the success of a harm reduction approach depends critically on when one examines outcomes – in Figure 2, if one looked at the periods immediately after the introduction of the harm reduction method, things seem to have gotten worse, as both consumption of the original addictive good and the consumption of the harm reduction method are high. However, after some time periods the steady-state consumption of the original addictive good goes below the level it would have in the scenario where the harm reduction method is not available.[15]

[15] The presence of oscillatory trajectories in the model implies that the short-run responses to the introduction of a harm reduction method may not accurately predict the long-run outcomes. More generally, even in the absence of oscillations, a discrepancy between short-run and long-run dynamics is a possible feature of dynamic models. For instance, in a simpler model, Dragone and Vanin (2022) show that price and income changes

Perhaps in recognition of the fact that the introduction of a harm reduction method can lead to a short-run increase in the consumption of the original addictive good, best practices for harm reduction state that there are circumstances under which MOUD is best provided in an inpatient setting (SAMHSA, 2021b). This is particularly true when the patient has a history of heavy/binge use of the addictive drug. In other words, certain patients may be at risk of relapsing with the original drug while initiating an addictive harm reduction method, and hospital admission may represent a pre-commitment to not relapse with the original addictive good.

4.2 When do harm reduction policies decrease or increase health harms?

In the previous sections, we have examined the conditions under which the introduction of a harm reduction method can increase or decrease consumption of the original addictive good. In this section, we examine the conditions under which it decreases or increases health harms, which may be an important outcome to policymakers and health care providers.

For individuals who were abstaining from the addictive good (e.g. non smokers, non opioid users), the answer is simple: if they begin using the harm reduction method, health harms worsen. The reason is that, even though the harm reduction method is less harmful than the original good, it still has some health harms. And for those previously abstaining, consumption of the harm reduction method cannot be accompanied by any reduction in the use of the original addictive good, so there are no offsetting reductions in health harms. Moreover, if the harm reduction method is highly addictive, it is possible that it acts as a gateway drug, inducing initiation of the addictive good (see Table 3). This is clearly the worst case scenario from the public health perspective, with dual addictive consumption and worse health among persons who were previously abstaining from the original addictive good.

Remark 5 *If a previous abstainer of the original addictive good uses the harm reduction method, health harm increases.*

We next consider people who were previously using the original addictive good. To examine the conditions under which harm reduction can decrease or increase health harms for this group, rewrite equation (13) in terms of long-run change in health harm after the introduction of a harm reduction method:

$$H_d^{ss} - H^{ss} = \theta_H \left(u_{vA} - r_L \left(u_{vc} \right) \right) v_d^{ss} \tag{16}$$

can have opposite effects in the short and long run when the law of motion of the state variable exhibits self-productivity.

Proposition 2 *After the introduction of a harm reduction method, health harms among users of the original addictive good eventually decrease if the harm reduction method is mildly addictive, and increase if the harm reduction method is moderately or highly addictive.*

The key factors determining whether health harms increase or decrease among this group are: 1) the addictiveness of the harm reduction method; and 2) the substitutability of the harm reduction method for the original addictive good.

Consider addictiveness. If the harm reduction method is mildly addictive, $u_{vA} < r_L(u_{vc})$, then the stock of health harms H falls (see Table 2). Moreover, as shown in the previous Section, for a previous consumer of the original addictive good c, the consumption of c also falls. This represents an unambiguous success of the harm reduction approach – introducing the new addictive option leads to a reduction in consumption of the original addictive good that is large enough to compensate for the health harm due to the use of the harm reduction method. Hence health ultimately improves (see Figure 2, left-most panel, for an illustration).

The steady state level of health harms can rise if the addictiveness of the harm reduction method is moderate (see eq. 16). The reason is that, even though consumption of the addictive good has declined, the individual is also using the harm reduction method, which itself contributes to both the addictive stock A and the stock of health harms H. If the consumption of the harm reduction method increases substantially, it is still possible for overall health to deteriorate. In one sense the harm reduction approach has been successful - it has reduced consumption of the original addictive good – but in another sense, it has failed because it has worsened the health of those who were previously using the addictive good.

Finally, if the harm reduction is highly addictive, not only does health harm increase, but so does the consumption of the original addictive good. In such a case, the harm reduction policy is unambiguously a failure. Thus, if the harm reduction method is highly addictive, then it backfires.

Note that threshold $r_L(u_{vc})$, like the threshold $r_H(u_{vc})$, is a function of the substitutability of the harm reduction method for the original addictive good. In particular, $\partial r_L/\partial u_{vc} < 0$ and the following Remark holds:

Remark 6 *The greater the substitutability of the harm reduction method for the original addictive good (i.e. the lower u_{vc}), the higher the threshold level $r_L(u_{vc})$.*

Hence, the greater the substitutability between the harm reduction method and the original addictive good, the more likely that the introduction of the harm reduction method will lead to a decrease in health harm. The intuition is straightforward: since the harm reduction method is less harmful than the original addictive good, the more the harm reduction method

is perceived by the consumer as a close substitute for the original addictive good, the more likely is the consumer to reduce consumption of c and increase consumption of v, resulting in a decline in health harms. In contrast, if the consumer perceives the harm reduction method as a complement with the original addictive good, that would make joint consumption of the two substances more likely, worsening health harms.

Table A1 provides an overview of the results for both abstainers and previous consumers of the addictive good. Note that, when examining the impact on health harms, timing is once again critical. If one examined only early time periods, one might miss later quitting of the original addictive good and perhaps even quitting of the harm reduction method. Immediate evaluation of harm reduction methods may give a misleading impression of steady-state effects.

Remark 7 *The greater the addictiveness of the original good, the less likely that health harms are reduced.*

As shown in the Appendix, both thresholds r_L and r_H decrease with u_{cA}. Intuitively, the higher the addictiveness of the original addictive good, the smaller the range of cases in which health harms are reduced.

5 Noteworthy special cases

5.1 Taxing the harm reduction method

In this Section we show how changes in the price of the harm reduction method due to, e.g. changes in taxation, can influence the demand for addictive consumption and health harm. (We show the effect of changes in the price of the harm reduction method here, and we show the effect of a change in the price of the original addictive good in Appendix A.4.)

Due to the absence of income effects, direct price effects are negative: $\partial v_d^{ss}/\partial p_v < 0$. For those who continue to consume the harm reduction method after its price has increased, the magnitude of the effect on Δc and ΔH becomes smaller. However, the signs of these effects remain unchanged because the threshold levels r_H and r_L do not depend on p_v (see equations 74 and 75). Specifically:

$$\frac{\partial c_d^{ss}}{\partial p_v} > 0 \qquad \Longleftrightarrow \qquad u_{vA} < r_H\left(u_{vc}\right) \tag{17}$$

That is, when the harm reduction method is mildly or moderately addictive, an increase in the price of the harm reduction method induces an increase in the consumption of the original addictive good. Empirically, the evidence shows that when the total price of vaping rises (e.g. due to taxes) then smoking of combustible cigarettes increases (see, e.g. Pesko and

Currie, 2019; Pesko et al., 2020; Pesko and Warman, 2022), which suggests that vaping is moderately or mildly addictive for consumers. As shown in the previous Section, these two cases correspond to the scenarios in which introducing the harm reduction method leads to a reduction in the consumption of the original addictive good and, possibly, a reduction in health harms. Under such conditions, taxing harm reduction methods may worsen health harms. A similar categorization is possible also for other harm reduction methods such as, e.g., Methadone, by observing the consumption patterns and behavioral responses of individuals when those methods become available.

The health consequences of higher taxes on the harm reduction method v can be explicitly assessed considering that

$$\frac{\partial H_d^{ss}}{\partial p_v} > 0 \qquad \Longleftrightarrow \qquad u_{vA} < r_L\left(u_{vc}\right) \tag{18}$$

That is, an increase in the price of the harm reduction method increases health harm if the harm reduction method is mildly addictive. Consistent with the predictions presented in the previous Section, we conclude that taxing the harm reduction method can produce different results, depending on its addictiveness.

Remark 8 *Taxing the harm reduction method:*

- *Increases consumption of the original addictive good if the harm reduction method is either mildly or moderately addictive,* $u_{vA} < r_H\left(u_{vc}\right)$
- *Increases health harm if the harm reduction method is mildly addictive,* $u_{vA} < r_L\left(u_{vc}\right)$
- *Decreases consumption of the original addictive good and decreases health harm if the harm reduction method is highly addictive,* $u_{vA} > r_H\left(u_{vc}\right)$

An implication of the above Remark is that, in the intermediate case in which the harm reduction method is moderately addictive, $u_{vA} \in (r_L, r_H))$, taxation of the harm reduction method will increase consumption of the original addictive good, and yet lead to a health improvement, because the health benefit of reduced vaping outweighs the increased harm from greater smoking. If the harm reduction method is highly addictive then we know from the earlier results summarized in Table 2 that the harm reduction method is backfiring, and causing people to actually consume more of the addictive good. In this case, raising taxes on the harm reduction method has the benefit of reducing consumption of the original addictive good and reducing health harms.

Rather than tax the harm reduction method, governments may instead to decide to subsidize it; at the extreme, offering it for free. Public health insurance plans often fully cover MOUD,

for example. When the individual chooses between the original addictive good and the harm reduction method, they take into account all costs of each, both monetary and non-monetary. If the harm reduction method is free, it simply makes it cheaper to consume the good. This is a special case regarding price but not fundamentally different.

5.2 The harm reduction method is an antagonist

Until now, we have assumed that the harm reduction method acts like an agonist, in that it binds with and activates the same receptors of the original drug ($\varepsilon_A > 0$), it is pleasurable (i.e. gives an euphoric effect, $u_v > 0$), and it is addictive ($u_{vA} > 0$). We now allow for the harm reduction method to instead act as an antagonist, like the opioid antagonists naloxone and naltrexone, or the alcohol antagonist Antabuse (disulfiram). Like agonists, antagonists bind with the same receptors of the original addictive drug. The critical difference is that antagonists block these receptors, reducing the pleasure of consumption the original addictive good ($u_{vc} < 0$). Moreover, antagonists are typically not enjoyable ($u_v = 0$) or addictive ($\varepsilon_A = u_{vA} = 0$).

By imposing the above restrictions in our model, it is easy to show that a previous abstainer of the addictive good would not use the antagonist harm reduction method (see eq. 14), as it does not provide positive marginal utility of consumption. In contrast, previous consumers of the original addictive good may demand an antagonist harm reduction method, which, under certain conditions can lead to quitting the original addictive good and then the harm reduction method. As shown in the previous Section, along this optimal trajectory, patterns of intermittent consumption are possible. Accordingly, there may be periods in which the antagonist harm reduction method is used, even if eventually the individual will not demand it in the long run.

Overall, the harm reduction method being an antagonist is an interesting special case of the existing model, and has implications for the thresholds r_H and r_L, but is not fundamentally different.

5.3 Banning the harm reduction method

So far our analysis has focused on the introduction of a harm reduction method, but the model can also be used to examine the reverse: when a harm reduction method is withdrawn from the market. An example of this is when, in 2022, the U.S. Food & Drug Administration (FDA) attempted to ban Juul from marketing its ENDS products in the United States, while cigarettes remained legal to sell (U.S. Food and Drug Administration, 2022).

As shown earlier, when the harm reduction method is not highly addictive (i.e. it is either moderately or mildly addictive), its use is negatively related to the consumption of the addictive

good (i.e. they are substitutes). As a result, if the harm reduction method is banned, the consumption of the original addictive good is expected to increase. If the harm reduction method is mildly addictive (as opposed to moderately or highly addictive), the increase in consumption of c is so large that health harm increases. However, if the harm reduction method is highly addictive, then c and v move together as if they were complements. As a result, when the harm reduction method is no longer available, consumption of the addictive good c decreases, and due to the reduced consumption of both goods, health harm decreases.

6 Extension: A Dual-Self Model of Costly Temptation

The economic rationale for taxes or regulations include, from the neoclassical perspective, a desire to address negative externalities, and from the behavioral economics perspective, a desire to address "internalities" arising from temptation and costly self-control (Herrnstein et al., 1993). In this section, we allow for the latter by incorporating a dual-self model, in which observed choices are the result of a disagreement between a rational, forward-looking self that takes into account long-term consequences, and a myopic, impulsive self that prefers immediate gratification without considering future consequences (Gul and Pesendorfer, 2004; Loewenstein and O'Donoghue, 2004; Fudenberg and Levine, 2006; Gul and Pesendorfer, 2007).

Consider a dual-self model in which one addictive good is available. The short-run self focuses on the instantaneous utility function $\mathcal{U}(c, q; A, H)$ and is fully myopic - all future consequences of current behavior are ignored. Let $\hat{c} = \hat{c}(A, H)$ be the *tempting choice*, i.e. the optimal choice for the short-run self.

In contrast, the long-run self maximizes intertemporal utility taking into account the future consequences of current behavior. Making the farsighted rather than myopic choice $\hat{c}$ imposes a utility cost, which can be interpreted as the cost of resisting temptation. We assume that this self-control cost is proportional to the amount of consumption foregone by resisting temptation. The long-run self's objective function is therefore

$$V = \int_0^\infty e^{-\rho t} \left[\mathcal{U}(c, q; A, H) - (\hat{c} - c) B\right] \mathrm{d}t \tag{19}$$

where $B \geq 0$ is a parameter influencing the intensity of the temptation cost.

The problem is solved by the long-run self, which chooses the consumption path that maximizes (19) subject to the law of motions (1) and (2), and the budget constraint (5). Despite featuring temptation and self-control costs, the solution to this problem is time-consistent. In other words, the consumer does not ever succumb to temptation, so there are not preference reversals as in, e.g. Gruber and Köszegi (2001), or regret as in, e.g. Orphanides and Zervos

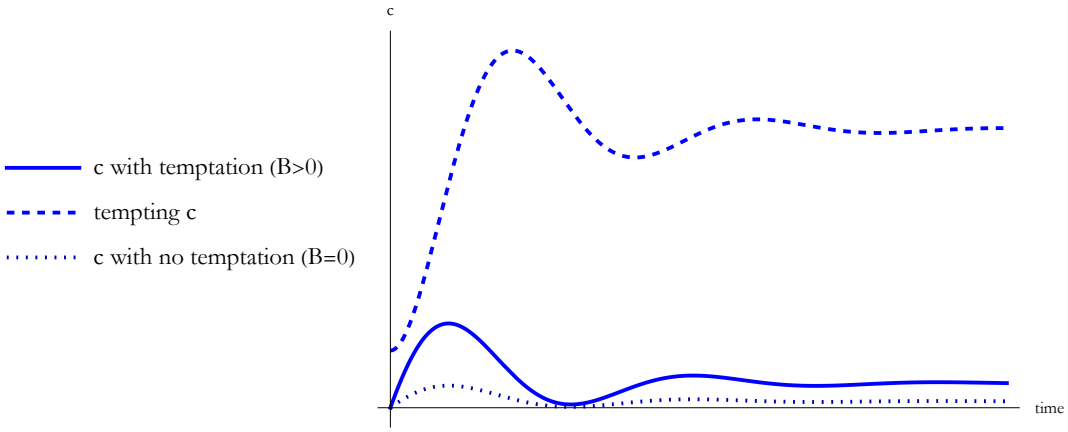

Figure 3: Trajectories of addictive consumption when there is only one addictive good. Tempting consumption, which is the preference of the impulsive myopic self (and indicated by the dashed line), actual consumption chosen by the farsighted self who faces temptation costs (solid line), and consumption in the absence of temptation (dotted line). Parameters as for Figure 1 and $B = 25$

(1995). Although the consumer doesn't succumb to temptation, they do pay a price for having to resist it, and this price is what can cause the optimal consumption path in this dual-self model to differ from that of the earlier model. If the agent's long-run preferences absent temptation are those relevant for social welfare maximization, then policy intervention is justifiable on the basis of the departure from that optimal consumption path caused by temptation costs (Schelling, 1984; O'Donoghue and Rabin, 1999, 2001, 2003).

6.1 Self-restraint and consumption under temptation costs

To assess some of the consequences of temptation costs, consider the case in which only the original addictive good is available. As shown in Appendix A.1, steady-state consumption in presence of temptation costs is

$$c^{ss} = c^{ss}_{B=0} - \beta \left[\left(\delta_A + \rho \right) u_{cc} + u_{cA} \right] B \tag{20}$$

where $c^{ss}_{B=0}$ denotes the steady state consumption in the benchmark case with no temptation, and $\beta > 0$. The second term of the right hand side of equation (20) shows how temptation costs cause consumption to deviate from the benchmark case.

Although one might expect that the presence of self-control costs leads to higher consumption of the addictive good, this is not a general result. In fact, $c^{ss} > c^{ss}_{B=0}$ only when the good is not sufficiently addictive, i.e. $u_{cA} < -\left(\delta_A + \rho \right) u_{cc}$. This case is illustrated in Figure 3, in which observed consumption (solid line) lies above the optimal consumption path absent temptation (dotted line). If instead the addictive good is sufficiently addictive, a forward-looking

consumer will exercise self-restraint, so that consumption of the addictive good will ultimately be lower in a scenario with temptation than in a scenario without temptation. Self-restraint may seem counter-intuitive in a model with both addiction and temptation costs, but a far-sighted individual would rationally seek to avoid accumulating excessive addiction, because they would foresee that it would result in substantial temptation costs in the future. Note that this depends crucially on farsightedness and rate of time discount.

When a person faces temptation, the introduction of a harm reduction method can have different effects on their consumption of the original addictive substance and the associated health harm, depending on the degree of addictiveness of the harm reduction method. This is similar to the benchmark case, with the difference that, as the temptation parameter B increases, the scope for achieving harm reduction broadens when considering self-control costs, and the set of cases where harm reduction may prove counterproductive backfire becomes smaller. Formally,

$$c_d^{ss} - c^{ss} > 0 \iff u_{vA} > r_H + \frac{\Omega_d}{\theta_d v_d^{ss}} B \tag{21}$$

$$H_d^{ss} - H^{ss} > 0 \iff u_{vA} > r_L + \frac{\Omega_d}{\theta_d v_d^{ss}} B \tag{22}$$

with $\theta_d, \Omega_d > 0$. Hence, as B increases, also the thresholds for consumption and health harm to increase become larger.

Figure 4 shows the time trajectories for tempting consumption, actual consumption and temptation costs, before and after the introduction of a harm reduction method. As temptation depends on all available alternatives, the temptation cost when the harm reduction method enters the market is given by $(\hat{c} - c + \hat{v} - v)B$, where $\hat{v} = \hat{v}(A, H)$ is the tempting choice of v. In the illustrative example shown in Figure 4, the harm reduction method is mildly reinforcing, temptation costs and consumption of the original addictive good decrease in the long run, and thus health harms are reduced.

6.2 Pigouvian-style taxation

If the agent's long-run preferences in the absence of temptation are relevant for social welfare maximization, the existence of temptation that induces choices to diverge from the long-run optimum justifies policy intervention (Schelling, 1984; O'Donoghue and Rabin, 2001, 2003). In other words, the intertemporal utility of an agent with no temptation cost $B = 0$, as implicitly assumed in the previous sections, can be taken as the normative reference by the social planner. Hence, policy actions can be justified to help consumers facing temptation costs $B > 0$ achieve the consumption path described in the previous sections.

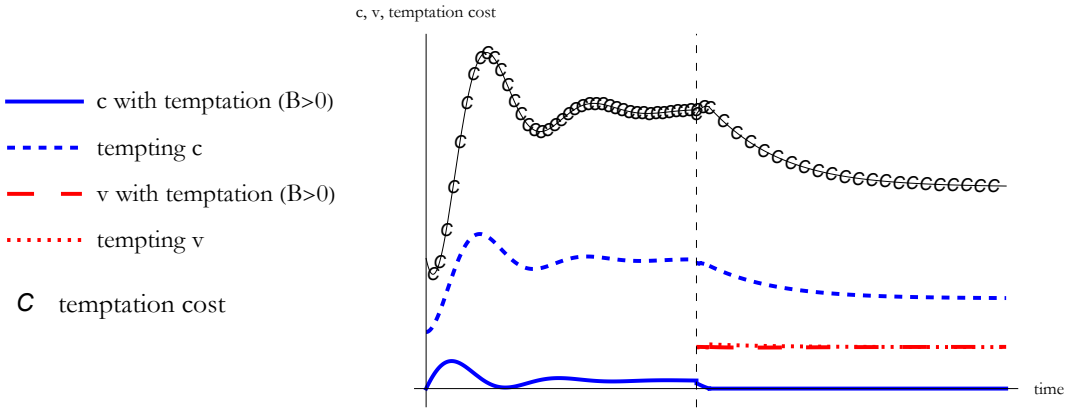

Figure 4: Tempting consumption, actual consumption and temptation costs, before and after the introduction of a harm reduction method. In this illustrative example, consumption of the original addictive good and temptation costs are lower after the introduction of the harm reduction method. Parameters as for the effective harm reduction scenario (mild addictiveness) of Figure 2 and $B = 2.3$

For example, a government might introduce a tax τ on the consumption of the addictive good. In a purely neoclassical model, a Pigouvian tax is designed to internalize an externality and thus incentivize consumption that maximizes social welfare. Similarly, in a behavioral economics model, a Pigouvian-style tax can be used to address an internality generated by temptation, and to induce the individual consumption trajectory to coincide with the one described in the previous sections, in which $B = 0$.

In case of only one addictive good, an excise tax that achieves this goal is

$$\tau = \left(1 + \frac{u_{cA}}{(\delta_A + \rho)u_{cc})}\right) B \tag{23}$$

which, consistent with equation (20), increases or decreases with the temptation parameter B depending on whether the original addictive good is sufficiently addictive, i.e. $(\delta_A + \rho)\, u_{cc} + u_{cA}$ is smaller or larger than zero.[16] Similar results hold when there exists not just an addictive good but also a harm reduction method, with Pigouvian-style taxes on each.

7 Discussion

Harm reduction methods are controversial. Advocates argue that they can increase the quitting of addictive substances, and, even if not, will reduce overall health harms. Opponents argue

[16] Due to the linearity of the temptation cost function, the Pigouvian-style tax does not depend on time, the state of addiction, or health harm. Moreover, it aligns the steady state and the *whole time path* of addictive consumption with the normative reference of addictive consumption absent temptation costs. Note also that, in principle, the tax on the addictive good τ could be negative, i.e. a subsidy, to counteract the situation in which an individual subject to temptation self-restrains consumption with respect to the benchmark case that the policy maker uses as the normative reference.

that they may backfire, make quitting less likely and leading to increased addiction, and that previous abstainers may initiate the use of the harm reduction method or even the original addictive good.

This paper outlines the conditions under which each of these predictions is correct. We provide a model of harm reduction, an implication of which is that the introduction of a novel harm reduction technique is neither always good nor always bad. Depending on the characteristics of the harm reduction method, it may not be consumed at all, may be consumed by those previously taking the original addictive good, and/or may be consumed even by those who previously abstained from the original addictive good. Also, depending on the characteristics of the harm reduction method, it can lead current users of the addictive good to quit, it can lead current users to increase their consumption of the original addictive good, or it can lead past abstainers to initiate the original addictive good.

There are three critical characteristics of the harm reduction method that determine which of these outcomes will occur. The first is its enjoyableness – do the benefits of the harm reduction method in terms of marginal utility of consumption exceed the costs in terms of monetary price and future health harms? This will determine whether people consume the harm reduction method. For those who do not consume it, nothing changes. They continue to have the same steady-state consumption of the original addictive good as before.

For those who do consume the harm reduction method, whether or not it leads previous users of the original addictive good to quit or not, and whether it leads previous abstainers to begin using the original addictive good, is determined by the second and third critical factors. The second factor is the addictiveness of the harm reduction method. This is critical because the more the harm reduction method contributes to the addictive stock, the more it increases the marginal utility not just of the harm reduction method but also the original addictive good. A harm reduction method that is highly addictive will be more likely to lead previous users of the addictive good to increase their consumption and will be more likely to induce previous abstainers to initiate the use of the addictive good. A government would likely not knowingly approve for sale a harm reduction method that it knew to be highly addictive, but its addictiveness may not be known with certainty at the time it is introduced. For example, after Bayer developed heroin, they marketed it for almost 20 years as a safe and less addictive substitute for morphine (Brunton and Knollmann, 2023). Thus, it is worth considering the case in which the harm reduction method turns out to be highly addictive.

The third critical factor is the extent to which the new harm reduction method is a substitute for (as opposed to a complement to) the original addictive good. The greater the extent to which it is a substitute, the less likely it leads previous abstainers to initiate and the less likely

it worsens health harms.

An important insight from the model is that the effect of the new harm reduction method depends critically on which time period is examined. Depending on the time period examined, one might see the use of original addictive good increasing or decreasing. In our simulations, the situation sometimes seems worse in the early periods, with use of both the harm reduction and original addictive good rising initially. Under some conditions, however, the outlook improves with time, as people decrease their consumption over time as health consequences mount. Thus, one should be careful that evaluations conducted immediately after the introduction of the harm reduction may be misleading, and it may take time to determine how the harm reduction method has affected steady-state consumption of the original addictive good.

The model also indicates that there are trade-offs to reducing access to harm reduction methods. On the one hand, restricting access to the harm reduction method can reduce the health harms that arise specifically from the harm reduction method, but on the other hand restricting access to the harm reduction method makes it harder for consumers to switch away from the original addictive good, potentially leaving them in worse health and more heavily addicted.

We show that the model can accommodate dual selves, in which the farsighted self pays a temptation cost to resist the desires of the myopic self. Reducing temptation costs for the farsighted self (who takes into account long-term consequences) represents an economic rationale for government intervention, which in this case could take the form of approving harm reduction methods for sale, or regulation of its characteristics.

The model of harm reduction used in this paper applies to a variety of cases, including ENDS and NRT for combustible cigarettes, methadone and buprenorphine for heroin and other opioids, and edible THC products for combustible marijuana. It applies to not only agonist therapies which have some euphoric and addictive aspects, but also to antagonist therapies which not only have no euphoric properties but reduce the marginal utility of the original addictive good.

The model implies a variety of policy levers that the government can use to affect the likelihood that the introduction of the harm reduction method succeeds in reducing health harms and consumption of the original addictive good:

1) Whether the government chooses to allow the harm reduction method on the market at all. For example, a government may decide whether to give regulatory approval for a new prescription drug, such as buprenorphine, or a new over-the-counter product such as Electronic Nicotine Delivery Systems (ENDS) or e-cigarettes. Currently, 32 nations ban e-cigarettes. (WHO, 2022d). In the U.S., the FDA authorized the marketing of ENDS devices in 2021 (U.S.

Food and Drug Administration, 2021) but the same agency a year later issued marketing denial orders to Juul, to prohibit them from selling their ENDS products in the U.S. (U.S. Food and Drug Administration, 2022).

2) The government can selectively regulate access to the harm reduction method. Governments may in particular want to restrict access by youth (Abouk and Adams, 2017; DeSimone et al., 2022). In 2019 the U.S. raised the minimum age to purchase e-cigarettes from 18 to 21 (U.S. Food and Drug Administration, 2020b). It may also impose limits on a doctor's ability to prescribe prescription harm reduction products; for example, the U.S. historically limited the number of patients to whom a physician may prescribe buprenorphine (SAMHSA, 2021a). The government may require that these prescription methods be administered under certain conditions; for example, in the U.S., methadone is often provided only in a clinic; it is rarely given to patients for home consumption, as it is in the U.K., Canada, and Australia. In addition, harm reduction methods can be made available only with a prescription, rather than over-the-counter. This may help ensure that the harm reduction methods are consumed only by those most likely to benefit from them, and not those whose use might lead to worse health harms and greater addiction.

3) The government may regulate the addictiveness of the harm reduction method (u_{vA}). This factor turns out to be critical in determining what happens to the consumption of the original addictive good. Such regulation could, for example, limit the potency of buprenorphine or methadone doses, the amount of THC in edible marijuana products, and the amount of nicotine that can be delivered in an increment of time by an e-cigarette.

4) Governments may seek to reduce the health harms of the harm reduction method (ε_H in our model). In the model of this paper, we have taken the health harms of the harm reduction method as given, but there are ways that governments can affect this, such as setting high safety requirements for e-cigarette devices.

5) Governments may seek to decrease the marginal utility u_v of the harm reduction method. For example, in 2020, the FDA banned flavored ENDS that might appeal to youth, including fruit and mint flavors (U.S. Food and Drug Administration, 2020a) . Advertising may be seen as a complement to consumption of the advertised good (Becker and Murphy, 1993), so regulation of advertising may also reduce the marginal utility of the harm reduction method.

6) Governments may tax either the original good or the harm reduction methods in order to raise their monetary price and decrease demand for them (Pesko et al., 2020). There is substantial variation in the rate at which U.S. states tax e-cigarettes; 21 states do not tax them at all, and among states that do tax them the rates vary from 8% to 95% (IGEN, 2022). Clean indoor air laws that ban vaping also increase the total cost of vaping by raising its time cost.

7) Alternatively, governments may set the price of the harm reduction method at zero for those who are heavy users of the original addictive good. For example, government-subsidized drug treatment, such as MOUD, may involve giving people who use drugs zero-cost methadone in order to encourage them to decrease their use of heroin.

A limitation of the paper is that, while the model does yield equations for the steady-state consumption of the harm reduction method and the original addictive good, as well as the consumption paths leading to the steady state, there are difficulties in empirically estimating them because one cannot easily measure or observe key parameters such as the addictiveness of the two substances, the health harms of the two substances, and the marginal utility of the two substances. However, the model is useful for illuminating the factors critical to determining whether the introduction of a harm reduction method succeeds in reducing health harms and consumption of the original addictive good.

Despite its limitations, this paper contributes to the literature by proposing a model of harm reduction, the implications of which indicate that neither advocates nor opponents are always correct. The introduction of a harm reduction method can facilitate quitting and reduce health harms, as advocates claim, or can backfire and lead to not just to increased use of the addictive good and worsening health harms but the initiation of the addictive good by previous abstainers, as opponents fear. The model also indicates the key factors that determine which of these outcomes occur; these are the enjoyableness of the harm reduction method, the addictiveness of the harm reduction method, and the extent to which the harm reduction method is a substitute for (as opposed to a complement to) the original addictive good.

References

Abouk, R. and S. Adams (2017). Bans on electronic cigarette sales to minors and smoking among high school students. *Journal of Health Economics 54*, 17–24. 1, 7

Abouk, R., C. Courtemanche, D. Dave, B. Feng, A. S. Friedman, J. C. Maclean, M. F. Pesko, J. J. Sabia, and S. Safford (2023). Intended and unintended effects of e-cigarette taxes on youth tobacco use. *Journal of Health Economics 87*, 102720. 1

Abouk, R., R. L. Pacula, and D. Powell (2019). Association between state laws facilitating pharmacy distribution of naloxone and risk of fatal overdose. *JAMA Internal Medicine 179*(6), 805–811. 1

Allcott, H. and C. Rafkin (2022). Optimal regulation of e-cigarettes: Theory and evidence. *American Economic Journal: Economic Policy 14(4)*, 1–50. 1, 3.1

Alpert, A., W. N. Evans, E. M. Lieber, and D. Powell (2022). Origins of the opioid crisis and its enduring impacts. *The Quarterly Journal of Economics 137*(2), 1139–1179. 2

Arni, P., D. Dragone, L. Goette, and N. R. Ziebarth (2021). Biased health perceptions and risky health behaviors - Theory and evidence. *Journal of Health Economics 76*, 102425. 3.1

Auld, M. C. and P. Grootendorst (2004). An empirical analysis of milk addiction. *Journal of Health Economics 23*(6), 1117–1133. 3.1

Barrette, E., L. Dafny, and K. Shen (2021). Do policies to increase access to treatment for opioid use disorder work? NBER working paper 29001, National Bureau of Economic Research. 1

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy 76*(2), 169–217. 5

Becker, G. S., M. Grossman, and K. M. Murphy (1991). Rational addiction and the effect of price on consumption. *American Economic Review 81*(2), 237–241. 3.1

Becker, G. S., M. Grossman, and K. M. Murphy (1994). An empirical analysis of cigarette addiction. *American Economic Review 84*(3), 396–418. 8

Becker, G. S. and K. M. Murphy (1988). A theory of rational addiction. *The Journal of Political Economy 96*(4), 675–700. 1, 3, 3.1, 7

Becker, G. S. and K. M. Murphy (1993). A simple theory of advertising as a good or bad. *Quarterly Journal of Economics 108*(4), 941–964. 7

Brunton, L. L. and B. C. Knollmann (2023). *Goodman & Gilman's: The Pharmacological Basis of Therapeutics: A Textbook of Pharmacology, Toxicology, and Therapeutics for Physicians and Medical Students* (14th ed.). McGraw-Hill Education. Originally published in 1941 by Macmillan Publishing, New York. 7

Campbell, D. (2009). Vatican's stance on UN drugs policy "risks lives". https://www.theguardian.com/world/2009/feb/27/vatican-un-drugs-policy, last accessed on 2022-07-07. 3

Cawley, J. and C. J. Ruhm (2012). The economics of risky health hehaviors. In *Handbook of Health Economics*, Volume 2, pp. 95–199. Elsevier. 1, 3.1

Chaloupka, F. J. (1991). Rational addictive behavior and cigarette smoking. *Journal of Political Economy 99*(4), 722–742. 3.1, 8

Clarke, E., K. Thompson, S. Weaver, J. Thompson, and G. O'Connell (2019). Snus: A compelling harm reduction alternative to cigarettes. *Harm Reduction Journal 16*, 62. 2

Cotti, C., C. Courtemanche, J. C. Maclean, E. Nesson, M. F. Pesko, and N. W. Tefft (2022). The effects of e-cigarette taxes on e-cigarette prices and tobacco product sales: Evidence from retail panel data. *Journal of Health Economics 86*, 102676. 1

Darden, M. and D. Gilleskie (2016). The effects of parental health shocks on adult offspring smoking behavior and self-assessed health. *Health Economics 25*(8), 939–954. 3.1

DeSimone, J. S., D. S. Grossman, and N. R. Ziebarth (2022). Regression discontinuity evidence on the effectiveness of the minimum legal e-cigarette purchasing age. Working Paper 30614, National Bureau of Economic Research. 7

Dockner, E. J. and G. Feichtinger (1993). Cyclical consumption patterns and rational addiction. *American Economic Review 83*(1), 256–263. 3

Doleac, J. L. and A. Mukherjee (2022). The effects of naloxone access laws on opioid abuse, mortality, and crime. *The Journal of Law and Economics 65*(2), 211–238. 1, 2

Dragone, D. and D. Raggi (2018). Testing rational addiction: When lifetime is uncertain, one lag is enough. Working paper 1119, Department of Economics, Università di Bologna. 1, 3.1, 7, 3.1

Dragone, D. and D. Raggi (2021). Resolving the milk addiction paradox. *Journal of Health Economics 77*, 102452. 1, 3.1, 7, 3.1

Dragone, D. and P. Vanin (2022). Substitution effects in intertemporal problems. *American Economic Journal: Microeconomics 14*(3), 791–809. 6, 15

Dragone, D. and N. R. Ziebarth (2017). Non-separable time preferences, novelty consumption and body weight: Theory and evidence from the East German transition to capitalism. *Journal of Health Economics 51*, 41–65. 1

Ehrlich, I. and G. S. Becker (1972). Market insurance, self-insurance, and self-protection. *Journal of Political Economy 80*, 623–648. 2

Einav, L. and A. Finkelstein (2018). Moral hazard in health insurance: What we know and how we know it. *Journal of the European Economic Association 16(4)*, 957–982. 2

Erickson, P. G. (1995). Harm reduction: What it is and is not. *Drug and Alcohol Review 14*(3), 283–285. 1

Friedman, A. S. (2015). How does electronic cigarette access affect adolescent smoking? *Journal of Health Economics 44*, 300–308. 1

Frio, G. S. and M. T. A. França (2021). Human papillomavirus vaccine and risky sexual behavior: Regression discontinuity design evidence from Brazil. *Economics & Human Biology 40*, 100946. 1

Fudenberg, D. and D. K. Levine (2006). A dual-self model of impulse control. *American economic review 96*(5), 1449–1476. 1, 6

Grimm, C. A. (2020). *Geographic disparities affect access to buprenorphine services for opioid use disorder*. US Department of Health and Human Services - Office of the Inspector General, OEI-12-17-00240. 2

Gruber, J. and B. Köszegi (2001). Is addiction "rational"? Theory and evidence. *The Quarterly Journal of Economics 116*(4), 1261–1303. 3.1, 6

Gul, F. and W. Pesendorfer (2004). Self-control and the theory of consumption. *Econometrica 72*(1), 119–158. 1, 6

Gul, F. and W. Pesendorfer (2007). Harmful addiction. *The Review of Economic Studies 74*(1), 147–172. 1, 6

Harm Reduction International (2022). What is harm reduction? https://www.hri.global/what-is-harm-reduction, last accessed on 2022-07-07. 1

Herrnstein, R. J., G. F. Loewenstein, D. Prelec, and W. Vaughan Jr (1993). Utility maximization and melioration: Internalities in individual choice. *Journal of Behavioral Decision Making 6*(3), 149–185. 6

IGEN (2022). Vape & e-cig taxes by State. https://igentax.com/vape-tax-state, last accessed on 2022-07-07. 2, 7

Laporte, A., A. R. Dass, and B. S. Ferguson (2017). Is the rational addiction model inherently impossible to estimate? *Journal of Health Economics 54*, 161–175. 3.1

Lillard, D. R. (2020). *The Economics of Nicotine Consumption, in: K. F. Zimmermann (ed.), Handbook of Labor, Human Resources and Population Economics*. Springer. 2

Loewenstein, G. and T. O'Donoghue (2004). Animal spirits: Affective and deliberative processes in economic behavior. Working paper 04-14, Center for Analytic Economics, Cornell University. 1, 6

Maclean, J. C., J. Mallatt, C. J. Ruhm, and K. Simon (2021). Economic studies on the opioid crisis: A review. NBER working paper 28067, National Bureau of Economic Research. 1, 2

Margolis, J., J. Hockenberry, M. Grossman, and S.-Y. Chou (2014). Moral hazard and less invasive medical treatment for coronary artery disease: The case of cigarette smoking. NBER working paper 28067, National Bureau of Economic Research. 1

Marshall, A. (1890). *Principles of Economics*. Macmillan and Co. Ltd.: London. 2

Marti, J., J. Buckell, J. C. Maclean, and J. Sindelar (2019). To "vape" or smoke? Experimental evidence on adult smokers. *Economic Inquiry 57*(1), 705–725. 1

Mas-Colell, A., M. D. Whinston, J. R. Green, et al. (1995). *Microeconomic Theory*. Oxford University Press, New York. 7

National Harm Reduction Coalition (2023). Harm reduction issues: Medication for opioid use disorder. https://harmreduction.org/issues/medication-for-opioid-use-disorder/, last accessed on 2023-01-16. 1

O'Donoghue, T. and M. Rabin (1999). Doing it now or later. *American economic review 89*(1), 103–124. 6

O'Donoghue, T. and M. Rabin (2001). Choice and procrastination. *The Quarterly Journal of Economics 116*(1), 121–160. 6, 6.2

O'Donoghue, T. and M. Rabin (2003). Studying optimal paternalism, illustrated by a model of sin taxes. *American Economic Review 93*(2), 186–191. 6, 6.2

Orphanides, A. and D. Zervos (1995). Rational addiction with learning and regret. *Journal of Political Economy 103*(4), 739–758. 6

Packham, A. (2022). Syringe exchange programs and harm reduction: New evidence in the wake of the opioid epidemic. *Journal of Public Economics 215*(1), 104733. 2

Peltzman, S. (1975). The effects of automobile safety regulation. *Journal of Political Economy 83*(4), 677–725. 2

Pesko, M. F., C. J. Courtemanche, and J. C. Maclean (2020). The effects of traditional cigarette and e-cigarette tax rates on adult tobacco product use. *Journal of Risk and Uncertainty 60*(3), 229–258. 1, 5.1, 7

Pesko, M. F. and J. M. Currie (2019). E-cigarette minimum legal sale age laws and traditional cigarette use among rural pregnant teenagers. *Journal of Health Economics 66*, 71–90. 1, 5.1

Pesko, M. F. and C. Warman (2022). Re-exploring the early relationship between teenage cigarette and e-cigarette use using price and tax changes. *Health Economics 31*(1), 137–153. 1, 5.1

Piccoli, L. and S. Tiezzi (2021). Rational addiction and time-consistency: An empirical test. *Journal of Health Economics 80*, 102546. 3.1

Powell, D., R. L. Pacula, and E. Taylor (2020). How increasing medical access to opioids contributes to the opioid epidemic: Evidence from Medicare Part D. *Journal of Health Economics 71*, 102286. 2

Prochaska, J. J., E. A. Vogel, and N. Benowitz (2021). Nicotine delivery and cigarette equivalents from vaping a JUUL pod. *Tobacco Control 31*(e1), e88–e93. 3.2

Rees, D. I., J. J. Sabia, L. M. Argys, D. Dave, and J. Latshaw (2019). With a little help from my friends: The effects of Good Samaritan and naloxone access laws on opioid-related deaths. *The Journal of Law and Economics 62*(1), 1–27. 1, 2

SAMHSA (2021a). FAQs about the new buprenorphine practice guidelines. https://www.samhsa.gov/medication-assisted-treatment/become-buprenorphine-waivered-practitioner/new-practice-guidelines-faqs, last accessed on 2022-07-07. 7

SAMHSA (2021b). *Treatment Improvement Protocol (TIP) 63, Medications for Opioid Use Disorder*. Rockville, MD: U.S. Department of Health and Human Services. Publication No. PEP21-02-01-002. 4.1

SAMHSA (2023a). Harm reduction. https://www.samhsa.gov/find-help/harm-reduction, last accessed on 2023-01-18. 2

SAMHSA (2023b). Harm Reduction Framework. https://www.samhsa.gov/find-help/harm-reduction/framework, last accessed on 2024-06-04. 1

SAMHSA (2023c). Waiver Elimination (MAT Act). https://www.samhsa.gov/medications-substance-use-disorders/waiver-elimination-mat-act, last accessed on 2023-10-30. 2

Samuelson, P. A. (1965). A catenary turnpike theorem involving consumption and the golden rule. *American Economic Review 55*(3), 486–496. 7

Satel, S. (2019). The vaping overreaction. The Atlantic, https://www.theatlantic.com/ideas/archive/2019/10/danger-vaping-bans/600451, last accessed on 2022-07-07. 3

Schelling, T. C. (1984). Self-command in practice, in policy, and in a theory of rational choice. *The American Economic Review 74*(2), 1–11. 6, 6.2

Simon, K., A. Soni, and J. Cawley (2017). The impact of health insurance on preventive care and health behaviors: Evidence from the first two years of the ACA medicaid expansions. *Journal of Policy Analysis and Management 36*(2), 390–417. 1

Singer, J. A. and S. Hamilton (2023). Expand access to methadone treatment: Remove barriers to primary care practitioners prescribing methadone. *CATO Policy Analysis 960*, 1–17. 2, 3

Single, E. (1995). Defining harm reduction. *Drug and Alcohol Review 14*(3), 287–290. 1

Smart, R., D. Powell, R. L. Pacula, E. D. Peet, R. Abouk, and C. S. Davis (2023). Investigating the complexity of Naloxone distribution: Which policies matter for pharmacies and potential recipients. NBER Working Papers 31142, National Bureau of Economic Research. 1

Smith, V. K., D. H. Taylor Jr, F. A. Sloan, F. R. Johnson, and W. H. Desvousges (2001). Do smokers respond to health shocks? *Review of Economics and Statistics 83*(4), 675–687. 3.1

Stancliff, S., H. Joseph, C. Fong, T. Furst, S. D. Comer, and P. Roux (2012). Opioid maintenance treatment as a harm reduction tool for opioid-dependent individuals in NYC: The need to expand access to buprenorphine in marginalized populations. *Journal of Addictive Diseases 31*(3), 278–287. 2

Toxvaerd, F. (2019). Rational disinhibition and externalities in prevention. *International Economic Review 60*(4), 1737–1755. 1

U.S. Department of Health and Human Services (2023). Harm reduction. https://www.hhs.gov/overdose-prevention/harm-reduction,last accessed 2023-18-01. 2

U.S. Food and Drug Administration (2020a). FDA finalizes enforcement policy on unauthorized flavored cartridge-based e-cigarettes that appeal to children, including fruit and mint. FDA News Release, https://www.fda.gov/news-events/press-announcements/fda-finalizes-enforcement-policy-unauthorized-flavored-cartridge-based-e-cigarettes-appeal- last accessed on 2022-07-07. 7

U.S. Food and Drug Administration (2020b). Newly signed legislation raises federal minimum age of sale of tobacco products to 21. https://www.fda.gov/tobacco-products/ctp-newsroom/newly-signed-legislation-raises-federal-minimum-age-sale-tobacco-products-21, last accessed on 2022-07-07. 7

U.S. Food and Drug Administration (2021). FDA permits marketing of e-cigarette products, marking first authorization of its kind by the agency. FDA Press release. https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-e-cigarette-products-marking-first-authorization-its-kind-agency, last accessed on 2022-07-07. 7

U.S. Food and Drug Administration (2022). FDA denies authorization to market JUUL products. FDA Press release. https://www.fda.gov/news-events/press-announcements/fda-denies-authorization-market-juul-products, last accessed on 2022-07-07. 2, 4, 5.3, 7

Vestal, C. (2016). In opioid epidemic, prejudice persists against methadone. Pew Charitable Trusts, https://www.theatlantic.com/ideas/archive/2019/10/danger-vaping-bans/600451, last accessed on 2022-07-07. 3

Waters, R. C. (2019). In support of buprenorphine deregulation. *Family Practice Management 26*(4), 6–8. 2

Weber, E. and A. Gupta (2019). *State Medicaid programs should follow the 'Medicare Model': Remove prior authorization requirements for buprenorphine and other medications to treat opioid use disorders*. Legal Action Center, New York, NY. 2

WHO (2022a). Alcohol. https://www.who.int/news-room/fact-sheets/detail/alcohol, last accessed on 2022-07-07. 1

WHO (2022b). Opioid overdose. https://www.who.int/news-room/fact-sheets/detail/opioid-overdose, last accessed on 2022-07-07. 1

WHO (2022c). Tobacco. https://www.who.int/health-topics/tobacco, last accessed on 2022-07-07. 1

WHO (2022d). Tobacco: E-cigarettes. https://www.who.int/news-room/questions-and-answers/item/tobacco-e-cigarettes, last accessed on 2022-07-07. 2, 7

A Appendix

A.1 Solving the benchmark model with one addictive good

Unless otherwise noted, all proofs concern the case with temptation costs. The results presented in the absence of temptation costs correspond to the case where $B = 0$.

Given the objective function (19) and the laws of motion (1) and (2), construct the Hamiltonian function associated with the consumer's problem by substituting the value of q from the budget constraint:

$$\begin{aligned}\mathcal{H}(c, v; A, H) = &\left(u_c + \frac{u_{cc}}{2}c + u_{cA}A\right)c + M - p_c c - (\hat{c} - c)B \\ &+ \left(u_A + \frac{u_{AA}}{2}A\right)A + \left(u_H + \frac{u_{HH}}{2}H\right)H + u_{AH}AH \\ &+ \lambda(c - \delta_A A) + \mu(c + \omega A - \delta_H H)\end{aligned} \tag{24}$$

where λ and μ are the shadow prices of A and H, respectively, and

$$\hat{c} = \frac{u_c - p_c + u_{cA}A}{-u_{cc}} \tag{25}$$

is the tempting choice. Note that the temptation cost

$$(\hat{c} - c)B = \left(\frac{u_c + u_{cA}A - p_c}{-u_{cc}} - c\right)B \tag{26}$$

is increasing in the enjoyableness u_c of the addictive good and in the addiction stock A, and decreasing in c and in its price p_c.

The necessary conditions for an internal solution are

$$\mathcal{H}_c = 0 \quad \Leftrightarrow \quad \underbrace{u_c + u_{cc}c + u_{cA}A}_{\mathcal{U}_c} = p_c - \lambda - \mu - B \tag{27}$$

$$\dot{\lambda} = (\rho + \delta_A)\lambda - \omega\mu - \underbrace{\left[u_A + u_{AA}A + u_{AH}H + \left(\frac{B}{u_{cc}} + c\right)u_{cA}\right]}_{\mathcal{U}_A} \tag{28}$$

$$\dot{\mu} = (\rho + \delta_H)\mu - \underbrace{(u_H + u_{HH}H + u_{AH}A)}_{\mathcal{U}_H} \tag{29}$$

$$\dot{A} = c - \delta_A A \tag{30}$$

$$\dot{H} = c + \omega A - \delta_H H \tag{31}$$

together with the appropriate initial and transversality conditions. The first-order condition (27) implies that the marginal benefit of consuming c must be equal to the marginal cost of

consuming, which depends on the market price as well as on the shadow price of A and H, and on the temptation parameter B. Note that the addiction stock affects the consumption of the original good c directly (through $u_{cA}A$) and indirectly (through its shadow value λ), while health harms play only an indirect role through μ. The equation of motion of the shadow value of addiction (eq. 28) depends also on shadow value μ of health. Moreover, the marginal utility of addiction $\mathcal{U}_A$ directly depends on the addictiveness of c. In particular, u_{cA} reduces the shadow price of building up addiction because it increases the marginal utility of consuming the addictive good. The law of motion of the shadow value of health harms (eq. 29), instead, does not depend on addiction nor on c.

Solving (27) for c yields the optimal consumption of the addictive good:

$$c^* = \frac{u_c - p_c + B + u_{cA}A + \lambda + \mu}{-u_{cc}} \tag{32}$$

Replacing in (27) to (31) and imposing $\dot{\lambda} = \dot{\mu} = \dot{A} = \dot{H} = 0$ yields the steady state values λ^{ss}, μ^{ss}, A^{ss} and H^{ss}. Replacing these values in (32) yields

$$c^{ss} = c^{ss}_{B=0} - \beta\left[\left(\delta_A + \rho\right)u_{cc} + u_{cA}\right]B \tag{33}$$

$$H^{ss} = \frac{\delta_A + \omega_A}{\delta_A \delta_H} c^{ss}, \qquad A^{ss} = \frac{1}{\delta_A} c^{ss} \tag{34}$$

where

$$c^{ss}_{B=0} = \alpha\left(u_c - p_c - \pi_c\right) \tag{35}$$

and

$$\alpha = \frac{\delta_A \delta_H \left(\delta_A + \rho\right)\left(\delta_H + \rho\right)}{-u_{cc}|J|} > 0 \tag{36}$$

$$\beta = \frac{1}{-u_{cc}\left(\delta_A + \rho\right)}\alpha > 0 \tag{37}$$

$$\pi_c = -\frac{u_A}{\delta_A + \rho} - \frac{\delta_A + \rho + \omega_A}{\left(\delta_A + \rho\right)\left(\delta_H + \rho\right)} u_H > 0, \tag{38}$$

and $|J|$ is the determinant of the Jacobian matrix (not shown) computed at the steady state.[17] Recall that $u_{cc} < 0$ and observe that π_c decreases with u_A, u_H and ρ.

Note that, the first-order condition (27) computed at the steady state implies

$$\mathcal{U}_c + B = p_c - \lambda^{ss} - \mu^{ss} \tag{39}$$

[17] We assume that the trajectories to the steady state are asymptotically stable, which implies that we focus on the case in which two eigenvalues of the 4x4 Jacobian matrix associated with the dynamic system have non-positive real parts. When this is the case, $|J|$ is positive.

In steady state, λ and μ are negative, hence the right-hand side is positive. This allows for the possibility that $\mathcal{U}_c$ is negative, i.e. $c^{ss} > \hat{c}$. Specifically:

$$c^{ss} > \hat{c} \text{ if } (u_{cA} + \delta_A u_{cc}) \{[u_{cA} + (\delta_A + \rho) u_{cc}] \beta B - c^{ss}_{B=0}\} > \delta_A (u_c - p_c) \tag{40}$$

To derive the policy function to the steady state, replace (27) into the system of differential equations (28)–(31), and then solve the system for given boundary conditions A_0, H_0, A^{ss} and H^{ss}. The solution is a function of time, the initial conditions and a set of four eigenvalues and eigenvectors. Out of four eigenvalues, at least two have always positive real parts. Imposing asymptotic stability and replacing the two expressions that depend on time, yields the policy function,

$$\tilde{c}(A, H) = a_c c^{ss} + a_A A + a_H H \tag{41}$$

where

$$a_c = \frac{e_1 e_2}{\delta_A \delta_H} > 0 \tag{42}$$

$$a_A = \frac{(\delta_A + e_1)(\delta_A + e_2) + (\delta_A + \delta_H + e_1 + e_2)\omega}{\delta_A - \delta_H + \omega} \tag{43}$$

$$a_H = -\frac{(\delta_H + e_1)(\delta_H + e_2)}{\delta_A - \delta_H + \omega} \tag{44}$$

and e_1, e_2 are the eigenvalues with negative real parts associated to the Jacobian matrix of (28) to (31). If they are complex numbers, the coefficients of the policy functions are real, yet the policy function features oscillations. To see it, suppose $e_1 = z + yi$ and $e_2 = z - yi$, with $z < 0$ and $z, y \in \mathbb{R}$, then the coefficients are

$$a_c = \frac{z^2 + y^2}{\delta_A \delta_H} \in \mathbb{R} \tag{45}$$

$$a_A = -\frac{(z + \delta_z)^2 + y^2}{\delta_H - \delta_A - \omega} - \frac{2z + \delta_A + \delta_H}{\delta_H - \delta_A - \omega}\omega \in \mathbb{R} \tag{46}$$

$$a_H = \frac{(z + \delta_H)^2 + y^2}{\delta_H - \delta_A - \omega} \in \mathbb{R} \tag{47}$$

Since the Jacobian matrix does not depend on p_c, neither do the eigenvalues. Hence the price of c only affects (41) through changes in the steady state consumption level c^{ss}.

A.2 Pigouvian-style taxation

The Pigouvian-style tax τ is found by equating c^{ss} when $B > 0$ and the price is $p_c + \tau$, with $c^{ss}_{B=0}$. Using (33) and (35) the former corresponds to

$$c^{ss}_{\tau} = \alpha \left(u_c - p_c - \tau - \pi_c\right) u_{cc} - \beta \left[\left(\delta_A + \rho\right) u_{cc} + u_{cA}\right] B \tag{48}$$

and the latter corresponds to

$$c^{ss}_{B=0} = \alpha \left(u_c - p_c - \pi_c\right) u_{cc} \tag{49}$$

Equating the two expressions yields

$$\tau = \frac{\left(\delta_A + \rho\right) u_{cc} + u_{cA}}{u_{cc}\left(\delta_A + \rho\right)} B \tag{50}$$

Since the price of addictive consumption (which possibly includes the tax) only affects the position of the steady-state consumption, i.e. the intercept of the policy function (41), it is straightforward to conclude that, when an individual with temptation costs is subject to the Pigouvian-style tax τ on addictive consumption, the individual is induced to follow the same path of consumption, addiction and health harm that would be followed in the absence of temptation.

A.3 Two goods

The solution follows the same procedure used in the previous Section. The Hamiltonian function associated to the consumer's problem is

$$\begin{aligned}
\mathcal{H}\left(c, v; A, H\right) = & \left(u_c + \frac{u_{cc}}{2} c + u_{cA} A\right) c + \left(u_v + \frac{u_{vv}}{2} v + u_{vc} c + u_{vA} A\right) v \\
& + \left(u_A + \frac{u_{AA}}{2} A\right) A + \left(u_H + \frac{u_{HH}}{2} H\right) H + u_{AH} AH + M - p_c c - p_v v \\
& - \left(\hat{c} - c + \hat{v} - v\right) B \qquad (51) \\
& + \lambda \left(c + \varepsilon_A v - \delta_A A\right) + \mu \left(c + \varepsilon_H v + \omega A - \delta_H H\right) \qquad (52)
\end{aligned}$$

where λ and μ are the shadow prices of A and H, respectively, and

$$\hat{c} = \frac{1}{u_{cc} u_{vv} - u_{vc}^2} \left[\left(u_v - p_v\right) u_{vc} - \left(u_c - p_c\right) u_{vv}\right] + \frac{A}{u_{cc} u_{vv} - u_{vc}^2} \left(u_{vA} u_{vc} - u_{cA} u_{vv}\right) \tag{53}$$

$$\hat{v} = \frac{1}{u_{cc} u_{vv} - u_{vc}^2} \left[\left(u_c - p_c\right) u_{vc} - \left(u_v - p_v\right) u_{cc}\right] + \frac{A}{u_{cc} u_{vv} - u_{vc}^2} \left(u_{cA} u_{vc} - u_{vA} u_{cc}\right) \tag{54}$$

describe the static bliss points for c and v. If $u_{vc} > \max\{\frac{u_{cA}}{u_{vA}}u_{vv}, \frac{u_{vA}}{u_{cA}}u_{cc}\}$, then $\hat{c}$ and $\hat{v}$ are increasing in the addictive stock. Moreover, if $u_{vc} > \max\{u_{cc}, u_{vv}\}$, then $\hat{c} + \hat{v}$ is increasing in the enjoyableness of the two addictive goods (u_c and u_v), and in the addiction stock (A). In the proceeding, we assume these reasonable conditions hold. (This is indeed the case when $u_{vc} = 0$).

The necessary conditions for an internal solution are

$$\mathcal{H}_c = 0 \quad \Leftrightarrow \quad \underbrace{u_c + u_{cc}c + u_{cA}A + u_{vc}v}_{\mathcal{V}_c} = p_c - \lambda - \mu - B \tag{55}$$

$$\mathcal{H}_v = 0 \quad \Leftrightarrow \quad \underbrace{u_v + u_{vv}v + u_{vA}A + u_{vc}c}_{\mathcal{V}_v} = p_v - \lambda\varepsilon_A - \mu\varepsilon_H - B \tag{56}$$

$$\dot{\lambda} = (\rho + \delta_A)\lambda - \omega\mu - \underbrace{(u_A + u_{AA}A + u_{cA}c + u_{vA}v + u_{AH}H - B\zeta)}_{\mathcal{V}_A} \tag{57}$$

$$\dot{\mu} = (\rho + \delta_H)\mu - \underbrace{(u_H + u_{HH}H + u_{AH}A)}_{\mathcal{V}_H} \tag{58}$$

$$\dot{A} = c + \varepsilon_A v - \delta_A A \tag{59}$$

$$\dot{H} = c + \varepsilon_H v + \omega A - \delta_H H \tag{60}$$

together with the appropriate transversality conditions, and

$$\zeta = \frac{1}{u_{cc}u_{vv} - u_{vc}^2}\left[u_{vA}\left(u_{vc} - u_{cc}\right) + u_{cA}\left(u_{vc} - u_{vv}\right)\right] > 0 \tag{61}$$

The left-hand sides of the first order conditions (55) and (56) describe the instantaneous marginal utility of consuming c and v, respectively. The right-hand side of both expressions describes the marginal costs of consuming, which depend on the market price, on the temptation parameter, and on the shadow prices of A and H.

The equation of motion of the shadow value of addiction (eq. 57) depends also on the shadow value μ of health. Moreover, the marginal utility of addiction $\mathcal{V}_A$ directly depends on the addictiveness of c and v. In particular, u_{cA} and u_{vA} reduce the shadow price of building up addiction, due to the fact that addictiveness increases the marginal utility of consuming the addictive goods. The law of motion of the shadow value of health harms (eq. 58), instead, does not depend on addiction nor on c or v.

Solving the focs for c and v yields the optimal consumption of the addictive good and of the harm reduction method

$$c^* = a_1 u_{vc} - a_2 u_{vv}; \qquad v^* = a_2 u_{vc} - a_1 u_{cc} \tag{62}$$

where

$$a_1 = \frac{B + u_v - p_v + u_{vA}A + \varepsilon_A\lambda + \varepsilon_H\mu}{u_{cc}u_{vv} - u_{vc}^2} \tag{63}$$

$$a_2 = \frac{B + u_c - p_c + u_{cA}A + \lambda + \mu}{u_{cc}u_{vv} - u_{vc}^2} \tag{64}$$

In the special case in which the harm reduction method does not affect the marginal utility of the addictive good, i.e. $u_{vc} = 0$, c^* does not depend on the price or the marginal utility of the harm reduction method and, conversely, v^* does not depend on p_c, nor on the marginal utility of c.

Replacing c^* and v^* in (57) to (60) allows deriving the steady state values of λ, μ, A and H. Replacing in the expressions for c^* and v^* yields the steady-state consumption of the harm reduction method:

$$v_d^{ss} = \gamma c^{ss} + \theta_v\left(u_v - p_v - \pi_v\right) + B\Omega$$

where

$$\theta_v = \frac{-u_{cc}}{u_{cc}u_{vv} - u_{vc}}\frac{|J|}{|J_d|} > 0 \tag{65}$$

$$\pi_v = -\frac{\varepsilon_A}{\delta_A + \rho}u_A - \frac{\varepsilon_A\omega + \varepsilon_H\left(\delta_A + \rho\right)}{\left(\delta_A + \rho\right)\left(\delta_H + \rho\right)}u_H > 0 \tag{66}$$

$$\begin{aligned}\Omega = {} & \frac{\left(u_{vc} - u_{cc}\right)\left(u_{cA}u_{vc} - u_{vA}u_{cc}\right)}{u_{cc}\left(u_{cc}u_{vv} - u_{vc}^2\right)^2|J_d|}\left\{\delta_H\left(\delta_H + \rho\right)\left[u_{vA} + \delta_A u_{vc} - \varepsilon_A\left(u_{cA} + \delta_A u_{cc}\right)\right]\right. \\ & \left. - \left(\varepsilon_A - \varepsilon_H\right)\left[\delta_H u_{HA} + \left(\delta_A + \omega\right)u_{HH}\right]\right\} - \frac{\varepsilon_A u_{cA} + \left(\delta_A + \rho\right)u_{cc}}{\left(\delta_A + \rho\right)\left(u_{cc}u_{vv} - u_{vc}^2\right)}\frac{|J|}{|J_d|}\end{aligned} \tag{67}$$

$$\begin{aligned}\gamma = {} & \theta_v\left\{\underbrace{\frac{\varepsilon_A}{\delta_A + \rho}u_{cA} + \frac{1}{\delta_A}u_{vA}}_{>0} + u_{vc} + \underbrace{\frac{\varepsilon_A}{\delta_A\left(\delta_A + \rho\right)}u_{AA} + \frac{\left(\delta_A + \omega\right)\left[\varepsilon_A\omega + \varepsilon_H\left(\delta_A + \rho\right)\right]}{\delta_A\delta_H\left(\delta_A + \rho\right)\left(\delta_H + \rho\right)}u_{HH}}_{<0}\right. \\ & \left. + \underbrace{\left[\frac{\varepsilon_A}{\delta_H\left(\delta_A + \rho\right)} + \frac{\varepsilon_H}{\delta_A\left(\delta_H + \rho\right)} + \frac{\varepsilon_A\left(2\delta_H + \rho\right)\omega}{\delta_A\delta_H\left(\delta_A + \rho\right)\left(\delta_H + \rho\right)}\right]u_{AH}}_{<0}\right\}\end{aligned} \tag{68}$$

and $|J_d|$ is the determinant of the Jacobian matrix at the steady state (not shown).

Conditional on $v_2^{ss} \geq 0$, steady state consumption of the original addictive good, health

harm and addiction are

$$c_d^{ss} = c^{ss} + \theta_d \left(u_{vA} - r_H\right) v_d^{ss} - B\Omega_d \tag{69}$$

$$H_d^{ss} = H^{ss} + \left[\theta_d \left(u_{vA} - r_L\right) v_d^{ss} - B\Omega_d\right] \frac{\delta_A + \omega}{\delta_A \delta_H} \tag{70}$$

$$A_d^{ss} = \left\{c_d^{ss} + \left[\varepsilon_A + \theta_d \left(u_{vA} - r_H\right)\right] v_d^{ss} - B\Omega_d\right\} \frac{1}{\delta_A} \tag{71}$$

where

$$\theta_d = -\frac{\delta_A \delta_H \left(\delta_H + \rho\right)}{u_{cc}|J|} > 0 \tag{72}$$

$$\Omega_d = \frac{\delta_A \delta_H \left(\delta_H + \rho\right)\left(u_{vc} - u_{cc}\right)\left(u_{cA} u_{vc} - u_{cc} u_{vA}\right)}{u_{cc}^2 \left(u_{cc} u_{vv} - u_{vc}^2\right)|J|} > 0 \tag{73}$$

$$\begin{aligned} r_H = &-\left(\delta_A + \rho\right) u_{vc} - \frac{\left(\delta_A + \rho\right)\varepsilon_A}{\delta_A} u_{cA} - \frac{\varepsilon_A}{\delta_A} u_{AA} - \frac{\left(\delta_A + \rho + \omega\right)\left(\omega\varepsilon_A + \delta_A \varepsilon_H\right)}{\delta_A \delta_H \left(\delta_H + \rho\right)} u_{HH} \\ &+ \left[\frac{\varepsilon_A + \varepsilon_H}{\delta_H + \rho} + \frac{\left(\rho + 2\omega\right)\varepsilon_A}{\delta_A \left(\delta_H + \rho\right)} + \frac{\rho\left(\omega\varepsilon_A + \delta_A \varepsilon_H\right)}{\delta_A \delta_H \left(\delta_H + \rho\right)}\right] u_{AH} \end{aligned} \tag{74}$$

$$r_L = r_H - \frac{\omega\varepsilon_A + \delta_A \varepsilon_H}{\theta_d \left(\delta_A + \omega\right)} < r_H \tag{75}$$

Note that r_L and r_H decrease with u_{vc} and u_{cA}. Moreover

$$\frac{\partial r_H}{\partial \rho} = -\frac{\varepsilon_A}{\delta_A} u_{cA} - u_{vc} + \frac{\delta_A - \delta_H + \omega}{\delta_A \left(\delta_H + \rho\right)^2}\left[\varepsilon_A u_{AH} + \frac{\delta_A \varepsilon_H + \varepsilon_A \omega}{\delta_H} u_{HH}\right] \tag{76}$$

Comparing the steady state values yields:

$$c_d^{ss} - c^{ss} \geq 0 \iff \theta_d \left(u_{vA} - r_H\right) v_d^{ss} - B\Omega_d \geq 0 \tag{77}$$

$$H_d^{ss} - H^{ss} \geq 0 \iff \theta_d \left(u_{vA} - r_L\right) v_d^{ss} - B\Omega_d \geq 0 \tag{78}$$

Note that v_d^{ss} and Ω_d depend on u_{vA}, and are complicated expressions. Hence an explicit solution for (77) and (78) is difficult to describe. However, suppose they are satisfied with equality by two critical values of addictiveness, call them u_{vA}^c and u_{vA}^H, respectively.

With no temptation $(B = 0)$, then $u_{vA}^c = r_H$ and $u_{vA}^H = r_L$, which explains definition (1) for a mildly, moderately or highly addictive harm reduction method. To assess how the threshold value u_{vA}^c changes when B increases, apply the implicit function theorem to equation (77) to obtain $\frac{\partial u_{vA}^c}{\partial B} = \frac{\Omega_d}{D}$ where

$$D = \theta_d \left\{v_d^{ss} + \left(u_{vA} - r_H\right) \frac{\partial v_d^{ss}}{\partial u_{vA}}\right\} + B \frac{u_{vc} - u_{cc}}{u_{vv} u_{cc} - u_{vc}^2} \tag{79}$$

At $u_{vA} = r_H$ and $B = 0$, then $D = \theta_d v_d^{ss}$. Hence $\frac{\partial u_{vA}^c}{\partial B} > 0$ if $v_d^{ss} > 0$. This implies that the threshold value u_{vA}^c becomes larger than r_H when B increases. The same conclusion holds when considering (78), u_{vA}^H and r_L.

The policy functions for c and v as function of the state of addiction and health harm can be assessed as described in the case with one addictive good. It can be shown that

$$\tilde{c}_d(A,H) = a_{cd} + a_{Ad}A + a_{Hd}H \tag{80}$$

$$\tilde{v}_d(A,H) = \varpi_{vd} + \varpi_{Ad}A + \varpi_{Hd}H \tag{81}$$

A.4 Taxation

The effect on steady-state consumption of a change in own price (direct price effect) is

$$\frac{\partial v_d^{ss}}{\partial p_v} = \frac{u_{cc}}{(u_{cc}u_{vv} - u_{vc}^2)} \frac{|J|}{|J_d|} < 0 \tag{82}$$

$$\frac{\partial c_d^{ss}}{\partial p_c} = \frac{u_{vv}}{(u_{cc}u_{vv} - u_{vc}^2)} \frac{|J_v|}{|J_d|} < 0 \tag{83}$$

where $|J_v|$ is the determinant of the Jacobian matrix when the original addictive good is not available and the harm reduction method is instead available. Under asymptotic stability of steady-state use of the harm reduction method, $|J_v| > 0$.

When considering cross-price effects, we obtain

$$\frac{\partial c_d^{ss}}{\partial p_v} = \frac{\delta_A \delta_H (\delta_H + \rho)}{(u_{cc}u_{vv} - u_{vc}^2)|J_d|} (r_H(u_{vc}) - u_{vA}) \tag{84}$$

$$\frac{\partial v_d^{ss}}{\partial p_c} = -\frac{\delta_A \delta_H (\delta_H + \rho)(\delta_A + \rho)}{(u_{cc}u_{vv} - u_{vc}^2)|J_d|} \gamma \tag{85}$$

Finally, the effect of a price change of the harm reduction method on health harm in the case of dual consumption is

$$\frac{\partial H_d^{ss}}{\partial p_v} = \frac{(\delta_A + \omega)(\delta_H + \rho)}{(u_{cc}u_{vv} - u_{vc}^2)|J_d|} (r_L(u_{vc}) - u_{vA}) \tag{86}$$

Outcome	Conditions	Intuition	Steady-state
Currently Consuming the Original Addictive Good ($c^{ss} > 0$)			
Irrelevant harm reduction method: consumption of the original addictive good does not change	$u_v \leq p_v + \pi_v - \frac{\gamma}{\theta_v} c^{ss}$	The harm reduction method is not appealing enough	$v_d^{ss} = 0$ c, H constant
Substitution: the harm reduction method replaces the original addictive good:	$u_v > p_v + \pi_v - \frac{\gamma}{\theta_v} c^{ss}$	The harm reduction method is appealing. If it is	$v_d^{ss} > 0$ $c \downarrow$
Quitting the original addiction may result; Health harm may decrease	(i) $u_{vA} < r_L$	(i) mildly addictive, health harm decreases	(i) $H \downarrow$
	(ii) $u_{vA} \in (r_L, r_H)$	(ii) moderately addictive, health harm increases	(ii) $H \uparrow$
Worst case scenario: Consumption of both addictive goods increases. Health harm increases	$u_v > p_v + \pi_v - \frac{\gamma}{\theta_v} c^{ss}$ $u_{vA} > r_H$	The harm reduction method is appealing and highly addictive	$v^{ss} > 0$ $c \uparrow$ $H \uparrow$
Currently Abstaining from Original Addictive Good ($c^{ss} = 0$)			
Irrelevant: No initiation, no gateway effect	$u_v \leq p_v + \pi_v$;	The harm reduction method is not appealing enough	$v_d^{ss} = 0$ $c_d^{ss} = H_d^{ss} = 0$
Initiation with the harm reduction method only; no initiation with the original addictive good	$u_v > p_v + \pi_v$ $u_{vA} < r_H$	The harm reduction method is appealing but not highly addictive	$v_d^{ss} > 0$ $c_d^{ss} = 0$ $H \uparrow$
Gateway effect: Initiation with both addictive goods	$u_v > p_v + \pi_v$ $u_{vA} > r_H$	The harm reduction method is appealing and highly addictive. It induces initiation also with the original addictive good	$v_d^{ss} > 0$ $c, H \uparrow$

Table A1: Possible changes in steady-state outcomes after the introduction of a harm reduction method.