ELSEVIER

A theory of natural addiction

Trenton G. Smith [a,*], Attila Tasnádi [b]

[a] *School of Economic Sciences, PO Box 646210, Washington State University, Pullman, WA 99164-6210, USA*
[b] *Department of Mathematics, Corvinus University of Budapest, Fővám tér 8, 1093 Budapest, Hungary*

Abstract

Economic theories of rational addiction aim to describe consumer behavior in the presence of habit-forming goods. We provide a biological foundation for this body of work by formally specifying conditions under which it is *optimal* to form a habit. We demonstrate the empirical validity of our thesis with an in-depth review and synthesis of the biomedical literature concerning the action of opiates in the mammalian brain and their effects on behavior. Our results lend credence to many of the unconventional behavioral assumptions employed by theories of rational addiction, including adjacent complementarity and the importance of cues, attention, and self-control in determining the behavior of addicts. We offer evidence for the special case of the opiates that "harmful" addiction is the manifestation of a mismatch between behavioral algorithms encoded in the human genome and the expanded menu of choices faced by consumers in the modern world.
© 2006 Elsevier Inc. All rights reserved.

1. Introduction

The immature seed pod of *papaver somniferum*[1] contains a bitter, milky sap. Even in this, its most natural form, opium is a powerful drug, a stimulant narcotic poison that can induce hallucinations, profound sleep, or death. Reduced to its most sought-after chemical constituent,

[1] Commonly known as the opium poppy.

morphine, or further processed into heroin, opium is highly addictive and can have dramatic effects on the behavior, health, and well-being of its users. Opium's natural and synthetic derivatives (collectively known as the *opiates*) have well-known effects on human physiology and behavior: once they make their way into the bloodstream, opiates reliably induce a state of euphoria and pain relief, often followed by an increase in food consumption (Morley et al., 1985; Gosnell and Levine, 1996; McKim, 2002). Many who experience this state of mind find it pleasurable, and are inclined to try it again. But chronic use of opiates can result in severely impaired health,[2] and desperate addicts sometimes resort to theft or prostitution to obtain money to sustain the habit (National Institute on Drug Abuse, 2000). Given the potentially lamentable personal and social consequences of drug addiction (and the undeniable fact that legal restrictions have not been fully effective in eliminating drugs like heroin from the streets), many would agree that modern society would be much improved if our species could somehow rid itself of this particular human weakness.

Though the effects of opiates have been known to man for more than five millennia (Booth, 1996), only in recent decades has modern science made clear that opiate-like substances are also produced naturally in the bodies of humans and other animals. These substances are known collectively as the *endogenous opioids* and, like their poppy-derived counterparts, they have been shown to induce euphoria, pain relief, and appetite stimulation (van Ree et al., 1976; Yeomans and Gray, 1996; Mercer and Holder, 1997; Bodnar and Hadjimarkou, 2002).

The similarity of opiates and the endogenous opioids might seem something of a curiosity at first blush. Given the dramatic negative effects of opiates, what business do our bodies have producing their chemical cousins? There are, fortunately, many ways to answer this question, as the scientific literature is now replete with evidence demonstrating the circumstances under which our bodies produce endogenous opioids, the distribution of and variation in the endogenous opioid system across species, speculation about their evolutionary origins, and even confirmation that the biochemical "recipe" for endogenous opioids is firmly—and apparently universally—encoded in the human genome. This essay will attempt to identify circumstances under which a tendency to become "addicted" might serve a useful function, review supporting evidence from the biomedical literature, and ask what our findings might tell us about drug addiction. In other words, we will develop a theory of natural addiction.

2. Background

2.1. Rational addiction

A major source of inspiration for this investigation, and therefore a reasonable starting point for this essay, has been the rich body of theoretical and empirical work on addiction within the economics literature. This literature of *rational addiction* employs the formal mathematical tools of the economist in modeling addiction as a well-defined decision problem to be solved by an optimizing consumer. This approach allows for—and indeed, to some extent requires—the precise statement of the properties of the decision environment that generate addiction. It also

---

[2] The medical complications of chronic heroin use, for example, can include fetal death, scarred and/or collapsed veins, bacterial infections of the blood vessels and heart valves, abscesses (boils) and other soft tissue infections, disease of the liver or kidney, pneumonia, and tuberculosis. Death from overdose is not uncommon (National Institute on Drug Abuse, 2000).

allows for the application of the standard tools of welfare analysis in developing implications for drug policy.

The essential feature of most theories of rational addiction is the concept of *adjacent complementarity*, first employed in this context in 1988 by Becker and Murphy (1988). Adjacent complementarity requires that consumption of an addictive good today generates even more consumption of that good tomorrow—or more precisely, that the marginal utility of consumption increases with experience. This property is more general than the popular conception of addiction, of course, and—as Becker and Murphy emphasize—could be used to describe any consumptive behavior in which habits are formed.

The work of Becker and Murphy is notable for its bold assertion that the decision to consume addictive substances is indeed a decision, and as such it can be viewed as a rational decision in a standard economic framework: to be sure, the argument goes, there may be negative personal consequences stemming from addiction, but the fact that many people nevertheless choose to consume addictive substances suggests that—for these people—the benefits of addiction must outweigh the costs. From this beginning, behavioral implications such as the responsiveness of addicts (or potential addicts) to drug prices and criminal penalties, or the dynamics of addiction (e.g., why some people might choose to quit "cold turkey") can be derived. Indeed, the model offered by Becker and Murphy does seem to capture many aspects of the behavior of addicts, and its main empirical prediction (that announced increases in the *future* price of addictive goods should decrease *current* consumption) has been largely borne out in subsequent analysis (see, e.g., Becker et al., 1994; Grossman et al., 1998; and Gruber and Köszegi, 2001).

In spite of the success of the Becker–Murphy theory of rational addiction, several authors have subsequently noted that in many respects the particulars of Becker–Murphy are not consistent with what is known about the psychology of addiction and the subjective experience of addicts. It has been suggested, for example, that rather than the world of perfect information, foresight, and self-knowledge implicit in Becker–Murphy, addicts face uncertainty regarding the future consequences of addiction (Orphanides and Zervos, 1995, 1998), may have problems with self-control (Fehr and Zych, 1998; Gruber and Köszegi, 2001; Gul and Pesendorfer, 2001; O'Donoghue and Rabin, 2002), and may be influenced by emotional or psychological states (Loewenstein, 1996; Laibson, 2001; Bernheim and Rangel, 2002). There is clearly some truth in each of these critiques, but all of these authors continue to take as given the primitive behavioral property responsible for addiction: adjacent complementarity.[3] In what follows we take a step back from this descriptive approach and ask under what circumstances habit formation of the type implied by the theory of rational addiction might be optimal. In particular, given the universality of the brain chemistry that makes our species and many others susceptible to drug addiction, we ask under what *natural* conditions the quirky behavioral property known as adjacent complementarity might have arisen. It is our hope in doing so that a more parsimonious and richly descriptive theory will result.

---

[3] Though we emphasize complementarity here, precise formulations vary. Gul and Pesendorfer (2001), for example, define a good to be addictive if past consumption makes a person more prone to over-consume the good in the future; Bernheim and Rangel (2002) define an addictive good as one for which past consumption enables neutral cues to trigger a "hot state" that can lead a person to consume the good again even if additional consumption is not in his best interest; and Orphanides and Zervos (1998) generate intertemporal complementarity by allowing addictive goods to alter the time preferences of consumers. Nevertheless, these authors all rely on strong—or at least unconventional—primitive assumptions regarding the behavior of addicts.

2.2. A few words on biological foundations

Our aim, to be more explicit, is to identify circumstances in the evolutionary history of the human species in which addiction-like behavior was *optimal* in a well-defined sense. Because this approach is a departure from the standard practice among purveyors of economic theory,[4] our reasons for adopting it should be clarified. First, as outlined in the previous section, the pursuit of a psychologically realistic theory of rational addiction has generated a multiplicity of formal models—all of which claim some degree of generality—and it is not immediately obvious which should be applied to a given instance of habit formation. Second, the primitive behavioral assumptions in rational addiction theories vary widely, as do the corresponding implications for welfare analysis. The application of a naturalistic perspective to this *model selection problem* can help to resolve these two related shortcomings. Viewing habit formation from the perspective of behavioral biology can help to answer fundamental questions about, for instance, the role of information and uncertainty in decision-making, and can provide the investigator with well-defined conditions for the generation of "harmful" addictions.

In what follows, we will ultimately conclude that addiction is intimately and undeniably related to the phenomenon of associative learning, and that adjacent complementarity can arise from a simple Bayesian learning process. We will arrive at this conclusion not because it is the most intuitive explanation for substance abuse, nor because it is the explanation most consistent with the reported experience of addicts. Rather, we will argue that addiction-as-learning (or, under specified conditions, addiction-as-misplaced-learning) is fully consistent with the observed *behavior* of addicts, and that no other explanation is consistent with the evidence from neuroscience.

Because the architecture of the human nervous system is rarely invoked as a source of empirical evidence in economics,[5] we begin by offering a brief explanation for our emphasis on internal biochemical events as a starting point for a theory of natural addiction. One might imagine, after all, that the principles of behavioral biology could be applied to the phenomenon of addiction without reference to neuroscience. This would require nothing more than a search for examples of addiction-like behavior exhibited by animals in their natural habitat, and followed by the identification of reasons why—in *natural* settings—such behavior might have given its practitioners an edge, over the ages, in the currency of survival and reproduction. Hypotheses thus arrived at would then be subject to the usual scrutiny of scientific method: variation in the relevant environmental variables would be expected to generate corresponding variation in addiction-like behavior, both within and across species, and so forth. This approach suffers from at least two drawbacks:

---

[4] Several authors have argued that knowledge of human evolutionary history might help to inform economic theory. See, for example, Hirshleifer (1977, 1985), Rogers (1994), Bergstrom (1996), Robson (2001), Samuelson (2004), and Smith (2002).

[5] This may be changing, as evidenced by a growing body of research being published under the rubric of "neuroeconomics"; for a recent review, see Camerer et al. (2005). Though these authors often make passing reference to the fact that the human brain bears the mark of a system that evolved to solve specific adaptive problems, the bulk of the research being done at the interface of economics and neuroscience has been aimed at using the techniques of the neuroscientist (e.g., brain imaging) to prove (or disprove) the predictions of economic theory. Our approach, on the other hand, is to address the question of biological origins directly, and to formulate hypotheses consistent with what is known about the human nervous system.

(i) in spite of the rigorous debate in the behavioral and medical sciences over what exactly constitutes an addiction, no consensus has emerged[6]; and

(ii) it might turn out that the behavioral manifestations of "addiction" in natural settings bear very little resemblance to their modern counterparts in neuroscience laboratories and urban ghettos.

The approach we have chosen, suggested in the opening paragraphs of this essay, is to begin with an addictive *substance*, make note of the internal biochemical and physiological changes it induces in users, and search for examples of circumstances in which these same internal changes are observed in animals in their natural habitat. These circumstances will then presumably lead, as above, to hypotheses about the natural origins of addiction. This approach is possible, of course, only when scientific knowledge of the relevant internal molecular processes is in a relatively advanced state. In what follows we will make use of the fact that heroin, the quintessential example of an addictive substance, affects its victims by mimicking the endogenous opioids, one of the most thoroughly studied molecular systems in modern neuroscience. We acknowledge at the outset that this approach, with its narrow focus on a single class of substances, runs the risk of generating conclusions with only limited generality; this issue will be discussed further in Section 4.1.

2.3. The adaptive function of endogenous opioids

2.3.1. Opiates and opioids

It has long been known that rats, given the opportunity, will self-administer morphine to the point of addiction. Whether pushing a lever to trigger an intravenous injection or sipping from a dilute solution, opiate-using rodents exhibit all the symptoms of addiction seen in their human counterparts: active substance-seeking behavior, reinforcement, tolerance, and withdrawal (see, e.g., Headlee et al., 1955; Weeks, 1962; van Ree et al., 1976). In addition to being exceedingly convenient for the purposes of conducting experimental research on addiction, the fact that we share such a complex trait with a relatively distant cousin in the animal kingdom suggests strongly that there is something deeply innate and biological about drug addiction.

The specifics of the activity of morphine within the body have become known relatively recently. One of the more useful early innovations in opioid research has been the discovery of drugs that block or counteract the effects of the opiates. These drugs, known as *opioid antagonists*, often have opiate-like chemical structures and exhibit little or no interaction with non-opiate drugs. The theoretical underpinning to the action of opioid antagonists is that they interact with *opioid receptors* and compete with the opiate ligand.[7] In other words, when an opiate molecule (or more generally, an *opioid agonist*) enters the bloodstream, it circulates through the body until it comes into contact and binds with an opioid receptor, which is then activated.

---

[6] Indeed, one aim of this essay is to propose a meaningful definition of addiction. This issue is discussed further in Section 3.2.1.

[7] Receptors and ligands are the locks and keys, respectively, of biochemistry. Though the degree of specificity can vary, ligands typically serve as the body's messengers: by virtue of their unique physical and chemical properties, ligands have the ability to selectively activate their target receptors, often triggering physiological responses at the cellular level. The textbook example of a ligand/receptor system is insulin, secreted by the pancreas in response to high blood sugar and detected by receptors throughout the body, touching off a variety of compensatory processes that bring blood sugar back into the normal range. Common subcategories of ligands include hormones, peptides, and neurotransmitters (Nelson, 2000).

If a large number of opioid receptors are activated simultaneously (e.g., if the concentration of opiates in the bloodstream is high), this triggers the cascade of physiological and behavioral changes associated with opiate use. Opioid antagonists, on the other hand, prevent the action of the opiates, often by binding with (but *not* activating) the target receptors, thus physically blocking the opiates from taking effect (Cooper et al., 2002). Though the opioid receptor was for many years merely a hypothetical construct, in the early 1970s advances in biochemical assay technology enabled scientists to confirm that there was indeed an opiate-specific receptor, located in cells throughout the body (though particularly concentrated, as it turns out, in certain regions of the brain) with test-tube reactivities mirroring the pharmacological activity of opiates and their antagonists, while exhibiting no reactivity with other drugs (Simon et al., 1973; Pert and Snyder, 1973; Terenius, 1974).

Opioid antagonists have been invaluable tools for addiction research. Early studies showed that the opioid antagonists *naloxone* and *naltrexone* effectively attenuate the physiological and behavioral effects of morphine in rats and monkeys, and even induce symptoms of withdrawal in morphine-using subjects (Weeks and Collins, 1976; Harrigan and Downs, 1978; Killian et al., 1978). The subsequent approval of these drugs for use in humans has provided evidence of their effects on subjective experience (Griffiths and Balster, 1979). Naltrexone, for example, is known to effectively block the feeling of euphoria associated with heroin use, and for this reason it was once viewed as a promising treatment for heroin addiction. Unfortunately, the effects of opioid antagonists on drug self-administration are not straightforward: just as rats and monkeys have been known to *increase* consumption of morphine and heroin in response to naltrexone treatment (presumably to compensate for the reduction in hedonic effect), human addicts aware of the effects of naltrexone will often voluntarily discontinue treatment in order to once again experience the hedonic pleasures of heroin. For this reason, methadone (a mildly addictive opioid agonist) is often used for weaning addicts from heroin, although naltrexone is sometimes an effective tool under controlled (inpatient) conditions, as a surgical implantation, or with particularly motivated patients (Jaffe and Martin, 1990; Mello and Negus, 1996).

In spite of the enormous body of research into the intricacies of the workings of the endogenous opioid system in humans and other animals, very little attention has been paid to the question of natural origins we hope to address here. It is known, however, that opiate administration causes subjects to increase short-term food intake, to develop a preference for the location/place of administration, to become insensitive to pain, and to lose interest in sexual activity (van Ree et al., 1999, 2000). These observations, together with studies pointing to the centrality of endogenous opioids in ingestive behavior, palatability, and food cravings (Mercer and Holder, 1997; Yeomans and Gray, 2002), are consistent with an adaptive function for the endogenous opioids in guiding feeding behavior in natural settings.[8] The next section will examine more closely the role of the endogenous opioids in modulating feeding behavior, and provide a sketch of the kind of adaptive problem they seem to be designed to solve. A formal statement of this problem is provided in Section 3.

---

[8] The endocrinology of feeding is, of course, more complicated than this: many other molecular signals have been implicated in short-term feeding behavior, including serotonin, dopamine, neuropeptide Y, and cholescystokinin. These molecules are neglected here for the sake of brevity. Readers interested in the molecular complexities of short-term ingestive behavior are referred to the review by Cooper and Higgs (1994); the molecular basis of the long-term regulation of caloric intake—a different but related adaptive problem—is reviewed in Cummings and Schwartz (2003).

2.3.2. Endogenous opioids and ingestive behavior

There is widespread evidence pointing to a central role for endogenous opioids in the short-term regulation of food intake. Numerous studies have shown that rats, for instance, will eat less after being injected with opioid antagonists (e.g., Holtzman, 1974; Ostrowski et al., 1981; Sanger et al., 1983; Simpkins et al., 1985), and the reverse is true for morphine and other opioid agonists, which reliably generate an *increase* in short-term food intake (Martin et al., 1963; Rudski et al., 1992; Gosnell and Levine, 1996).[9] Similar responses have been observed in a wide variety of other foraging animals, from slugs (Kavaliers et al., 1985) and cockroaches (Kavaliers et al., 1987) to cats (Foster et al., 1981), pigs (Baldwin and Parrot, 1985), and humans (Cohen et al., 1985; Atkinson, 1982).

A widely held view in the scientific community posits that opioids mediate food intake by influencing the perceived palatability of foods (Yeomans and Gray, 1996; Mercer and Holder, 1997). In human subjects, opioid antagonists reduce both the hedonic ratings of palatable foods and the pleasantness ratings of palatable food odors, but do not reduce stated hunger ratings (Fantino et al., 1986; Yeomans and Wright, 1991; Drewnowski et al., 1992; Yeomans and Gray, 1996). If it is true, as the behavioral effects of opioid agonists and antagonists seem to suggest, that endogenous opioids in our brains cause food to taste good, then we might expect that good-tasting food causes our brains to release endogenous opioids. There is evidence that this is indeed true: for instance, the consumption of sweetened foods (but not bitter foods) causes an immediate release of β-endorphin in the brains and cerebrospinal fluid of rats (Dum et al., 1983; Yamamoto et al., 2000), and acute exposure to sweets induces reduced pain avoidance in rats and human infants, an effect that can be reversed with naltrexone (Blass, 1986; Blass et al., 1987; Blass and Hoffmeyer, 1991).

So we are presented with the following puzzle: When an individual eats a food containing sugar, he triggers a biochemical cascade that causes him to eat more of that particular food, irrespective of his immediate caloric needs. Why might such a system have evolved? In other words, what competitive advantage might be gained by foraging animals that exhibit such a preference for sweet foods?

The answer given by behavioral ecologists is derived from the distribution of sugar in nature. High concentrations of simple carbohydrates are found in natural settings only in ripe fruit, raw honey, and mother's milk, all of which reliably contain a host of valuable micronutrients (and, importantly, a dearth of toxins).[10] This suggests a simple role for endogenous opioids in an optimal foraging framework: when an environmental cue (such as the presence of sugar) indicates that a particular food is likely to have nutritional value, endogenous opioids are released in the brain, generating the appropriate behavioral response.

Of course, sugar is not the only cue omnivorous animals use in distinguishing beneficial foodstuffs from harmful or useless ingesta: the tongues of humans and other omnivorous mammals have compound-specific receptors not only for simple carbohydrates but also for sodium, glu-

---

[9] Although opioid antagonists were once viewed as a promising tool in the treatment of obesity, most studies have shown that they have little effect on feeding or body weight in the long term (Si et al., 1986).

[10] This distribution is not accidental: in a textbook example of coevolution, fruit-bearing plants rely on the services of foraging animals to disperse their seeds. This explains why fruit nearly always contains bitter or sour compounds and a green skin prior to maturation (i.e., before the seed is viable) and—when fully ripened—comes packaged not only with a brightly colored skin and edible, nutritionally valuable flesh, but also with non-digestible seeds or pits (see, e.g., Raven et al., 1999, pp. 546–551).

tamate,[11] and (perhaps) essential fatty acids—all of which could serve as nutritional cues in natural environments—and also for many toxic compounds, the dangers of which are perceived as a bitter taste, each of which appears to trigger an opioid response (Gilbertson and Kim, 2002; Sullivan, 2002). In social animals such as humans, sheep, and chickens, it is also known that *social* cues play an important role in dietary choice,[12] the response to which also appears to be opioid-mediated, as evidenced by the coincidence of autism (a developmental disorder characterized by both a specific deficit in the ability to read social cues (Baron-Cohen, 1995) and by deficiencies in the endogenous opioid system (Kalat, 1978; Gillberg and Lonnerholm, 1985; Campbell et al., 1988; Leboyer et al., 1988; Leboyer et al., 1990)) and abnormal dietary behavior in childhood (Raiten and Massaro, 1986). And one of the most important classes of cues for foraging animals—those that recall the *location* of valuable food sources—also constitute one of the most easily demonstrated effects of drugs such as morphine and heroine (Mucha et al., 1982; Mucha and Iversen, 1984).

Although the endogenous opioids no doubt have functions other than the regulation of short-term ingestive behavior, and short-term ingestive behavior is no doubt regulated by numerous other neuroendocrine processes in addition to the endogenous opioids, our aim here is to effect a simple demonstration of the way in which the solution to one particular adaptive problem might generate addiction-like behavior. In the next section we offer a formal model in which an environmental cue (the representative example of which is sugar) serves as an aid in the solution of a basic foraging problem: avoidance of micronutrient deficiency.

3. A model, with supporting evidence

3.1. Informative cues and the diet problem

3.1.1. A balanced diet

In what follows we present a stylized model of nutritional ecology with informative cues. A foraging animal ("agent") is faced with a menu of two foods, x and a, and must choose how much of each to consume, given the limited capacity m of his gut and (physical) food densities 1 and $1/p$, respectively. There is a single limiting micronutrient for which there is a critical threshold: if the agent does not consume k units of nutrient, he will die. Unfortunately, this implies that survival is by no means certain, as the nutrient concentrations in foods x and a are independent random variables, denoted C_x and C_a, with distribution functions F_x and F_a, respectively. Our agent can, under these circumstances, do no better than to minimize the odds of death by malnutrition.[13] Formally, the *balanced diet problem* can be stated as follows:

$$
\begin{aligned}
\max_{x,a} \quad & P(C_x x + C_a a \geqslant k) \\
\text{s.t.} \quad & x + pa \leqslant m, \\
& x, a \geqslant 0.
\end{aligned}
\tag{1}
$$

---

[11] *Glutamate* is a form of the amino acid *glutamine*, a molecular building block of protein. Glutamate is found in many natural foods; it is also the "G" in the flavor enhancer MSG.

[12] See review in Smith (2004).

[13] Or equivalently, to maximize his probability of survival. In accordance with the principles of evolution by natural selection, agents able to calculate, intuit, or otherwise implement the correct solution to this problem would presumably out-compete their more death-prone brethren, and come to dominate the population in the long run.

Given the inherent uncertainty in this decision problem, it is clear that a cue providing new information about the nutritional value of one of the foods might alter the outcome. We will consider such a signal by positing two (informational) states of the world: one in which no cue is present, as above, and one in which a "positive" cue is observed, implying that the concentration of the limiting micronutrient in good a (the "addictive" good) is given by the random variable $\widehat{C}_a$, with distribution function $\widehat{F}_a$. To distinguish between these two information states, we will refer to the *no-cue balanced diet problem* and the *positive-cue balanced diet problem*. For purposes of illustration, we restrict our attention initially to distribution functions having the form $F(c; \gamma) = c^\gamma$, where $\gamma > 0$, and in particular, to the following parameterization:

Case 1. "Cobb–Douglas Cue":

$$F_x(c_x) = c_x$$
$$F_a(c_a) = c_a^\beta$$
$$\widehat{F}_a(c_a) = c_a^{\hat{\beta}},$$

where β and $\hat{\beta}$ are parameters of the distribution functions such that $\hat{\beta} > \beta > 0$ (Fig. 1).

We are now ready to state the following proposition:

Proposition 1. *If the agent faces concentration distributions described by Case* 1 *and the solution* (x^*, a^*) *to the no-cue balanced diet problem is such that* $x^*, a^* > k$, *then in the solution* $(\hat{x}, \hat{a})$ *to the positive-cue balanced diet problem, his consumption of good* a *will be strictly greater,* $\hat{a} > a^*$.

All proofs are provided in Appendix A.[14,15]

It is also possible to make a more general statement about the conditions necessary for the cue to result in an increase in consumption of good a (relative to the no-cue optimum a^*), and doing so will provide a useful illustration of the intuition behind the problem. In Proposition 2, we consider the effect on the probability of survival of making a small (ε) movement along the budget line in the direction of increasing a. Such a movement, of course, simultaneously decreases the amount of food x consumed, so there are two distinct effects on the probability of survival, which we denote the *ε-benefit* (i.e., the increase in probability of survival attributable to the increase in a) and the *ε-loss* (i.e., the decrease in probability of survival attributable to the decrease in x).[16] To formulate our more general condition we impose the following assumption.

---

[14] The assumption that x^* and a^* are greater than k in Proposition 1, made for analytical convenience, deserves comment. In the naturalistic interpretation given here, this assumption is equivalent to assuming that in the absence of a positive cue, the agent will choose to consume enough of each good that *if the nutrient concentration in either of these goods were* 100%, then his consumption of that good alone would be enough to ensure his survival. Given the typically minuscule concentrations of micronutrients in natural foods, and the large amounts of food typically ingested by foraging animals (relative to the required quantity of micronutrients), we expect that this condition would rarely be violated in natural settings.

[15] In Appendix A we show that in Case 1, for $x, a > k$, the agent's behavior will be observationally equivalent to that of an agent maximizing a Cobb–Douglas utility function (Corollary 2). The properties of this class of utility functions are well known.

[16] Explicit definitions of ε-benefit and ε-loss are provided in Appendix A.
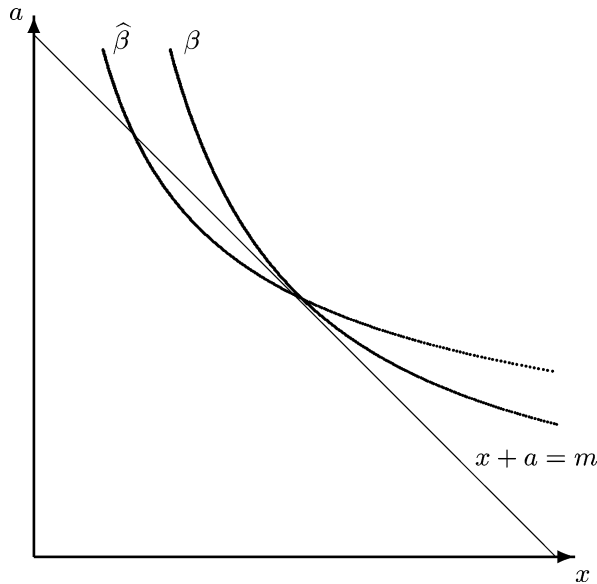
Fig. 1. Effect of a positive cue ($\beta = 1$ and $\hat{\beta} = 2$).

Assumption 1. The distribution functions F_x, F_a and $\widehat{F_a}$ have density functions f_x, f_a and $\hat{f}_a$, respectively.[17] Moreover, the probability mass is distributed over the entire unit interval, i.e., $f_x(c_x) > 0$ whenever $c_x \in (0, 1)$ and $f_x(c_x) = 0$ whenever $c_x \notin (0, 1)$. The corresponding conditions are also satisfied by f_a and $\hat{f}_a$.[18]

Now we are ready to state our necessary and sufficient condition.

Proposition 2. *Let Assumption 1 be fulfilled. In addition, assume that there exist unique solutions $x^*, a^* > k$ and $\hat{x}, \hat{a} > k$ to the no-cue balanced diet problem and the positive-cue balanced diet problem, respectively, and that the indifference curves associated with the objective functions of the no-cue balanced diet problem and of the positive-cue balanced diet problem going through point (x^*, a^*) cross only at point (x^*, a^*). Then a cue increases the consumption of a if and only if there exists an $\varepsilon > 0$ such that the ε-benefit exceeds the ε-loss.*

The intuition behind Proposition 2 is perhaps best illustrated by considering how various realizations of the random variables C_x and $\widehat{C}_a$ translate into survival. The probability of survival for a given allocation (x, a) can be represented graphically in (c_x, c_a) space. In Fig. 2, the line intersecting the vertical axis at $\frac{k}{a}$ (i.e., $c_x x + c_a a = k$) represents for our agent the ultimate threshold: any realization (c_x, c_a) above this line and the agent survives; any realization below and he dies. With an ε-movement along the budget line (i.e., to the allocation $(x - \varepsilon, a + \varepsilon/p)$), the threshold pivots to intersect the vertical axis at $\frac{k}{a+\varepsilon/p}$. This change in consumption increases

---

[17] We assume the existence of density functions to simplify the statement of Proposition 2.

[18] Restricting the support of these density functions to the unit interval is motivated by basic laws of physics: nutrient content cannot be less than zero or greater than 100%.
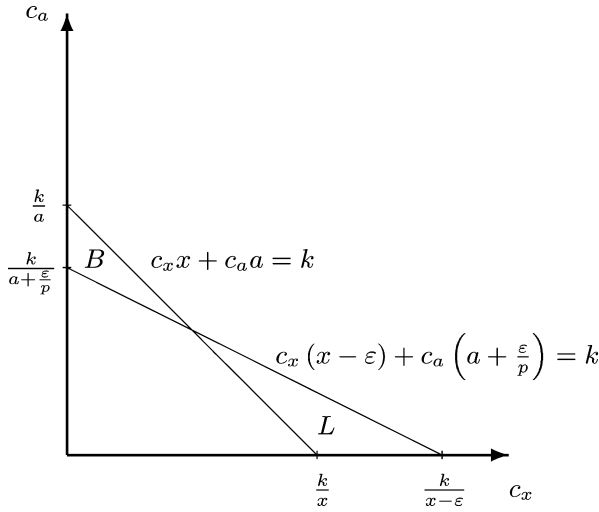
Fig. 2. ε-benefit and ε-loss.

the survival area by B while it decreases the survival area by L. Thus the probability that (c_x, c_a) falls in B is the ε-benefit, while the probability that (c_x, c_a) falls in L is the ε-loss.

Thus far we have addressed only the effect of an informationally valuable cue in a static choice environment. But if our aim is to say something about the relationship between this environment and addiction-like behavior, we will need to consider the dynamics of the balanced diet problem. In the next section we examine the simple dynamics of choice when learning is possible.

3.1.2. The simple dynamics of learning

In order to capture the notion of learning within our formal framework, it is necessary to re-formulate the decision problem slightly. In Section 3.1.1, consumption decisions were always made with complete knowledge of the probability distributions underlying the foods of choice. To accommodate learning, we will now impose a modicum of foresight on our agent, requiring him to choose a diet *before* confirming the presence or absence of a cue. In so doing, we will show how the simple dynamics of Bayesian learning can generate adjacent complementarity, the behavioral property driving theories of rational addiction.

Again, we distinguish between two states: the no-cue state in which the agent's objective function is given by $\bar{u}(x, a) = P(C_x x + C_a a \geqslant k)$ and the positive-cue state in which his objective function is given by $u(x, a) = P(C_x x + \widehat{C}_a a \geqslant k)$. The following assumption imposed on these two "utility" functions is consistent with the framework developed in Section 3.1.1.

Assumption 2. The functions u and $\bar{u}$ are twice continuously differentiable and strictly concave in the area satisfying $x, a > k$. In addition, the no-cue balanced diet problem and the positive-cue balanced diet problem have unique solutions $(x^{\bar{c}}, a^{\bar{c}})$ and (x^c, a^c) respectively, with $a^c > a^{\bar{c}}$; $x^c, a^c > k$; and $x^{\bar{c}}, a^{\bar{c}} > k$.

A bundle (x, a) maximizes the probability of survival in period t if it solves

$$\max_{x,a} \quad E\big[\Pi_t u(x, a) + (1 - \Pi_t)\bar{u}(x, a)\big]$$
$$\text{s.t.} \quad x + pa \leqslant m, \tag{2}$$
$$x, a \geqslant 0$$

where Π_t is the random variable describing the agent's prior beliefs in period t concerning the possible probabilities with which a positive cue might arise. Hence, Π_t maps into the space of probabilities and thus takes values on $[0, 1]$. We shall denote by g_t the density function describing the distribution of Π_t.[19]

In what follows, we write v_t for the objective function of problem (2). We can therefore rewrite problem (2) as $\max_{a \geqslant 0} v_t(m - pa, a)$ by the monotonicity of v_t. After choosing a bundle $(m - pa, a)$ the agent observes the period t outcome (i.e., presence or absence of a cue) and updates his beliefs for period $t + 1$ in a Bayesian manner. In order to isolate the effect of the positive cue, we will assume the agent's budget constraint remains the same in every period. If he observes a positive cue in period t, his posterior beliefs are given by

$$g_{t+1}(\pi) = \frac{\pi g_t(\pi)}{\int_0^1 \pi g_t(\pi)\, d\pi} \tag{3}$$

for all $\pi \in [0, 1]$. We shall denote by Π_{t+1} the random variable corresponding to density function g_{t+1}.

As noted in Section 2.1, the dynamic property known as adjacent complementarity has been identified as essential to a behavioral theory of addiction. In the present framework, an analogous property can be concisely defined:

Definition 1. The agent's behavior meets the conditions for *adjacent complementarity* at period t and at point $(m - pa, a) \in \mathbb{R}_+^2$ on the budget line if $\frac{d}{da} v_t(m - pa, a) < \frac{d}{da} v_{t+1}(m - pa, a)$.

Our next proposition states sufficient conditions for adjacent complementarity.

Proposition 3. *If Assumption 2 is satisfied, then a positive cue generates adjacent complementarity at any time t and at any point (x, a) lying on the budget line such that $a \in [a^{\bar{c}}, a^c]$. Moreover, if the agent chooses a bundle in period t such that $x^*, a^* > k$, then $a^* \in (a^{\bar{c}}, a^c)$ and $x^* = m - pa^*$.*

Corollary 1. *Under the assumptions of Proposition 3 the agent's behavior exhibits adjacent complementarity at the optimal solution (x^*, a^*) of problem (2).*

3.2. Addiction as learning gone awry

3.2.1. What is addiction?

In the theory of rational addiction, a good is addictive, roughly speaking, if its *marginal* (instantaneous) utility increases with experience. Rational addictions, however, may be either beneficial or harmful depending on how experience affects *total* (instantaneous) utility: in a beneficial addiction total utility increases over time, while in a harmful addiction it decreases over time. In the popular lexicon, the word addiction generally excludes the former case: one might have "good" as well as "bad" habits, but addiction is generally taken to imply a regrettable behavior. The careful reader will have noticed that thus far the model we have presented seems to imply that all "addiction" is beneficial: in the learning dynamic we have proposed, utility (or its proxy in our framework, the expected survival probability) is always increasing over time, as the

---

agent learns more about the world in which he lives. Now we turn our attention to the subject of harmful addiction.

The circumstances we propose as being conducive to harmful addictions are perhaps best illustrated by returning to our representative example of a behavioral cue. Although sugar is conveniently associated with valuable micronutrients in natural settings, the advent of commercially viable sugar refining technology early in the twentieth century changed this association dramatically. Today, foods with the highest sugar content often contain no micronutrients whatsoever; in fact, one of the most consistent messages of modern health advocates has been a simple admonition: *eat less sugar*. But the biochemical system upon which we rely in choosing our foods has not changed: being encoded in our genes in a way that leaves it mostly immune to conscious manipulation, the endogenous opioid system still reacts to sweet foods *as if* they remained a rare and valuable commodity. Within our framework, this implies a discrepancy between the behavior of the agent and maximization of the objective function v_t. We will find it useful, therefore, to specify a *subjective function*, $\tilde{v}_t$, that reflects the probability distributions $\widehat{F}_a$ and F_a and density function g_t prevalent in what might be called the agent's "ancestral environment"—that is, the conditions to which the agent's behavior is adapted.[20] This suggests the following definitions:

Definition 2. The agent's behavior meets the conditions for *subjective adjacent complementarity* at period t and at point $(m - pa, a) \in \mathbb{R}_+^2$ on the budget line if $\frac{\mathrm{d}}{\mathrm{d}a}\tilde{v}_t(m - pa, a) < \frac{\mathrm{d}}{\mathrm{d}a}\tilde{v}_{t+1}(m - pa, a)$.

Definition 3. The agent's behavior meets the conditions for a *harmful addiction* at the decision sequence $(x_1, a_1), (x_2, a_2), \ldots$ on the budget line if it satisfies subjective adjacent complementarity at (x_t, a_t) and $v_t(x_t, a_t) > v_{t+1}(x_{t+1}, a_{t+1})$ for all periods t.

As our example suggests, we will consider the special case in which the cue does not convey any information, so that $\widehat{F}_a$ is identical to F_a. This implies (for the *objective* function v) that $v_t = v_{t+1} = \bar{u}$ in equilibrium for all periods t independently of the agent's beliefs about the probabilities of the arrival of a positive cue. The agent's subjective function, however, still specifies that $\widehat{C}_a$ and C_a have different distributions. In Proposition 4 we denote by $(\tilde{x}_t, \tilde{a}_t)$ the solution of the problem that we obtain from (2) by replacing v_t with $\tilde{v}_t$.

Proposition 4. *Let Assumption 2 and $a^c > \tilde{a}_t > a^{\bar{c}}$ for all t be fulfilled. Moreover, suppose that although a dissociation of the cue from the limiting micronutrient results in an uninterrupted series of positive cues that provide no information, the agent continues to maximize the subjective function $\tilde{v}_t$. Then a harmful addiction will occur at the sequence $(\tilde{x}_1, \tilde{a}_1), (\tilde{x}_2, \tilde{a}_2), \ldots$.*

In effect, Proposition 4 says—for the special case of technological change considered here— that harmful addictions are the product of a mismatch between the modern world and the "beliefs" about the world implicit in our behavior.[21] This mismatch is generated, in the present

---

[20] Decision theorists have long held that beliefs can be thought of as *subjective*, or implicit in one's behavior. It is in this sense that we use the term. Leonard Savage's classic 1954 treatise provides an advanced but accessible exposition of this notion (Savage, 1954).

[21] The "evolutionary mismatch" theory of substance abuse represents the conventional wisdom among students of human evolution. See, for example, the work of Nesse and others (Williams and Nesse, 1991; Nesse, 1994; Nesse and Berridge, 1997 (but see also footnote 26)).

example, by the rapidity with which food processing technology has advanced while the human genome (in which, it bears repeating, the "recipe" for the endogenous opioid system is literally written) has remained effectively unchanged. In other words, our approach suggests that a harmful addiction is a habit acquired under false pretenses.

In spite of the evidence presented thus far, some readers may nevertheless remain uncomfortable with the notion of "sugar addiction". Recent studies by Hoebel and his colleagues have suggested that sugar does indeed share more properties with drugs of addiction than had previously been thought: feeding rats excessive amounts of sugar, for example, and then either depriving them of food or injecting naloxone induces symptoms typical of opiate withdrawal such as teeth chattering, forepaw tremor, and head shakes (Colantuoni et al., 2002). Remaining skeptics may find solace in the fact that replacing "sugar" with the word "alcohol" in this story will not change its character or consistency with available scientific knowledge.[22] The distribution of alcohol in nature mirrors that of sugar (i.e., it is found only in ripe fruit), it is subject to opioid-mediated self-administration,[23] and only industrial fermentation and distillation technologies have made it readily available in the modern world (Dudley, 2000). The recent identification of human genes that confer a higher risk of alcoholism provides further support for the notion that alcohol consumption might have had adaptive significance in human evolutionary history (Schuckit, 1999).

It does not seem inappropriate to suggest that heroin addiction also fits well with the "mismatch" model of addiction. When a user seeks out and injects heroin in order to experience once again the sudden activation of the opioid receptors in his brain, he is following an ancient algorithm: *when the opportunity arises, devote your energies to activities that make you feel like this*. The algorithm is, of course, more complicated than this, and vestiges of the original function of the behavior can be seen in some of the particulars of the experience of addicts: heroin addicts, for example, reportedly experience overpowering cravings for sugar during withdrawal (Weiss, 1982), and relapses among reformed addicts are often triggered by place-specific contextual cues (Carson-DeWitt, 2001). Heroin, like refined sugar and distilled alcohol, is a relatively recent innovation in our collective history, first synthesized from morphine in 1874.[24]

That the neurologically active substances such as morphine, cocaine, and nicotine found in plant tissues might be harmful to our health is not surprising when the origins of these compounds are considered. Once thought to be merely metabolic by-products, plant ecologists now believe that compounds like these (which are energetically costly to produce but have no apparent function in plant physiology) arose in the course of plant evolution as defensive mechanisms designed to deter herbivorous animals.[25] These substances effectively deter herbivory because they are highly potent neurotoxins: for instance, the oral ingestion of as little as 2 grams (0.07 oz.) of raw opium (or 0.2 grams refined morphine, or 0.04 grams nicotine), can be fatal (Parfitt, 1999). Coevolutionary forces work both ways, of course, and animals have in turn evolved methods of avoiding plant toxins, most notably by detecting and then ejecting them, either by tasting them directly with receptors on the tongue and spitting them out (bitter aversion) or by vomiting upon

---

[22] It is the experience of the authors that although few will acknowledge an overly zealous propensity for alcohol consumption in themselves, many are able to identify alcoholism in others.

[23] For a review of the large scientific literature implicating opioids in alcohol's addictive properties see Raven et al. (1999, pp. 375–378).

[24] In one of the more spectacular blunders in the annals of the pharmaceutical industry, heroin was originally developed and marketed by The Bayer Company as a less-addictive form of morphine (Booth, 1996).

[25] See, e.g., Raven et al. (1999, pp. 546–551).

the onset of illness (nausea aversion) (Smith, 2004). This helps to explain why drugs of addiction are commonly smoked, snorted, or injected but rarely chewed up and eaten: our bodies have natural mechanisms that prevent ingestion of toxins. That some of these toxins, taken in moderation, selectively activate specific "reward" centers in the brain that govern addiction appears to be an accident of plant-herbivore coevolution. This "accident" nevertheless displays all the hallmarks of an adaptation in natural settings.[26] Raw opium, for instance, contains a host of alkaloids in addition to morphine, many of which (e.g., papaverine, codeine, narcotine, and thebaine) have little or no narcotic effect but act as stimulants of the medulla and spinal cord (Parfitt, 1999). Taken together in their natural form, these compounds constitute a dangerous drug cocktail, and one important function of drug delivery technologies is to isolate—and thus detoxify—the target compound.

It has been conjectured that such a mismatch between objective reality and the "beliefs" about the world implicit in our genes could explain the subjective difficulties with "self-control" many people report when describing their experience with drugs of abuse or sweetened treats (Smith, 2002).[27] In the present context, the "belief" implicit in our genes (and also implicit—when viewed in light of the evidence from the natural sciences—in our behavior) is that foods containing simple carbohydrates are nearly always nutritionally valuable, while the objective reality is that in today's world such foods are more often than not lacking in such value. A literal interpretation of our theory of harmful addiction implies an agent who perpetually expects a large benefit to accrue from a particular activity, but—when the expected benefits are not realized—finds himself constantly regretting his past actions.[28] We contend that this interpretation has meaningful parallels with economic theories of self-control or "time inconsistent" behavior: in these theories, the agent typically overweights (i.e., assigns higher utilities to) current consumption to the detriment of his long-term well-being (i.e., his future utility stream) (Ainslie, 1991; Laibson, 1997). Within our framework, time inconsistency is the manifestation of emotional mechanisms maladapted to certain aspects of the modern world and underlain by molecular processes that operate below the level of consciousness.

3.2.2. Dopamine and the neurobiology of learning

Even the most devout hedonist would admit that habit formation (including drug addiction) can be viewed as a case of learning. If the consumer finds that consumption of a particular good gives him pleasure, for example, it would make sense to take this information into account when making subsequent consumption decisions. If subsequent consumption provides additional information (presumably also of a positive nature) about the hedonic properties of the good (e.g., higher levels of consumption correspond to more intense pleasure), then the resulting behavioral dynamic could closely approximate that predicted by a theory of rational addiction. But this explanation quickly stretches thin where harmful addictions are involved: the typical drug addict quickly becomes aware of the hedonic properties of his drug of choice, and his behavior often

---

[26] This hypothesis is not completely uncontroversial—some argue that substance abuse may have been around long enough for the human genome to have developed defensive mechanisms (see, e.g., Sullivan and Hagen, 2002). The debate, however, is mostly one of degree: no one would argue, for example, that hypodermic needles have been around long enough for humanity to develop an innate aversion to heroin.

[27] Indeed, in some situations admitting reasonable levels of subjective uncertainty can transform an apparent self-control problem into an optimal behavioral strategy. Sozou, for example, shows that hyperbolic discounting of future rewards can be optimal where default is possible and the hazard rate is uncertain (Sozou, 1998).

[28] This interpretation is, of course, *overly* literal, because our hypothetical "expectation" need not be conscious or even subject to conscious control.

becomes increasingly pathological (i.e., less informed by the dictates of informed rationality) over time. Is it really appropriate, as we have suggested, to view harmful addiction as some kind of learning disorder?

One way of answering this question would be to identify the neurological basis of learning: if our hypothesis is correct, drugs of abuse would be expected to act on the same physical substrates employed in healthy, natural learning processes. While modern science is far from providing a definitive picture of how the internal workings of the mammalian brain translate into sophisticated learning abilities, several authors have noted that one process in particular conforms well with the predictions of classical learning theory: dopamine transmission in the limbic system.[29,30,31] Although dopaminergic neurons are present in many areas of the mammalian brain, many of those located in limbic system appear to have the intriguing property of being subject to activation by natural rewards only when associative learning is taking place: in classic Pavlovian conditioning experiments, these neurons are activated *in the presence of novel, but not conditioned stimuli.* In other words, these neurons seem to indicate that *learning* is taking place: when an animal is first presented with a novel visual or auditory stimulus while being fed a tasty treat, dopaminergic neurons in his limbic system light up; but with experience, the stimulus/treat pairing loses the ability to activate these neurons. Once a subject is conditioned, only "surprises"—such as the pairing of food reward with a stimulus not previously associated with the reward—will reactivate the system.

So what are the effects of opiates on dopamine transmission in the limbic system? Interestingly, stimulation of dopamine transmission in this part of the brain is one of the few properties shared by virtually all drugs of addiction—not only opiates, but also alcohol, nicotine, cocaine, amphetamines, and Δ^9-tetrahydrocannabinol.[32] The most powerfully addictive drugs, however, differ from natural rewards in that their effects on dopamine transmission are not diminished by repeated administration.[33] This would seem to suggest that if dopamine in the limbic system does in fact represent a physical substrate of associative learning, then the sort of learning that takes place in the presence of drugs of addiction is properly viewed as pathological.[34]

---

[29] Though precise definitions vary, the limbic system in the mammalian brain is comprised of several interconnected structures, generally including the amygdala, hippocampus, hypothalamus, septum, nucleus accumbens, cingulated gyrus, and parts of the cortex. The limbic system is thought to play a central role in the regulation of emotions (Brick and Erickson, 1998).

[30] Although neurons (nerve cells) can employ more than one neurotransmitter at a given synapse (a *synapse* is the gap between cells across which neurotransmitter ligands carry information), much intercellular communication in the mammalian brain is mediated by neurotransmitters such as dopamine, norepinephrine, epinephrine, or serotonin in distinct cells. Hence the terms "dopaminergic neuron," "serotinergic neuron," etc. See, for example, Zigmond et al. (1999).

[31] Several authors have focused particular attention on the nucleus accumbens shell and the ventral tegmental area as the putative loci of associative learning (Chiara, 1999; Spanagel and Weiss, 1999). Unfortunately, brain mapping is far from an exact science, and similar observations have been made in other brain structures. The advanced state of current technology, however, is evident in the recent report of Waeltl et al. (2001)—complete with simultaneous measurement of eye position and the activity of individual neurons in the subjects' brains—in which monkeys were trained to associate the delivery of fruit juice with distinctive visual stimuli.

[32] Δ^9-tetrahydrocannabinol, or THC, is the pharmacologically active constituent of marijuana.

[33] The difference presumably stems from the fact that the endogenous neurochemical signals generated by natural rewards are subject to adaptive regulation; exogenous ligands (i.e., drugs) are not subject to such limitations. It is important to note that the distinction is not absolute: drugs of addiction *are* subject to habituation, but to a much lesser degree than natural rewards.

[34] It is worth noting that the problem of "attention" emphasized in some alternative theories of rational addiction (e.g., Laibson, 2001; Bernheim and Rangel, 2002) also appears to be a function of endogenous opioid activity, as evidenced by

4. Discussion

4.1. Beyond opioids

From the discussion of the previous section, it is clear that the logic of our analysis might well apply to substances or circumstances that activate receptor systems having nothing to do with the opioids. We urge caution on those who would extend our analysis in this way; and though for the remainder of this essay we will turn to a more general discussion of habit formation, we do so with some degree of trepidation. That some drugs target other receptor systems suggests that they are disrupting a different adaptive response. In principle, before drawing conclusions about the adaptive function of the receptor system targeted by a given drug of abuse, systematic study of the role of the target system in natural settings should be undertaken. Once the adaptive function of the target receptors are identified, specific implications (including, for example, circumstances under which it is likely to be used, or the sense in which it meets the criteria for "harmful" addiction) for the relevant addictive substance would presumably follow. The prospect of such an undertaking for all behavior-altering substances is daunting, given the unfortunate fact that—for most drugs—our understanding of the myriad effects of such substances on human behavior, cognition, and physiology remains poor.

4.2. Foreseeing addiction

As noted in Section 2.1, the question of the degree to which consumers foresee the consequences of drug addiction is the subject of much debate in the rational addiction literature. Though we have intentionally suppressed the possibility of foresight in our formal analysis (in order to emphasize that even myopic decision-making can generate the dynamic properties necessary for habit formation), there can be no doubt that consumers are to some extent aware of the future (social, health, and economic) consequences of drug abuse. Not that people "know" the consequences of addiction in the same way people "know" it is good to eat ripe fruit, or the way people "know" the diet that sustained them in childhood is unlikely to harm them later in life. But there's every reason to expect that people could learn from health that worsens with use, from watching relatives die or suffer, from reading about health consequences, or from warnings on labels. Indeed, such learning could well be interpreted within our framework as constituting "informative cues" that influence the agent's beliefs (i.e., $\widehat{F}_a$, F_a, and g_t) and behavior accordingly. Or, extending our framework a bit, an agent who becomes aware of the harmful dynamic associated with drug use might well choose to quit "cold turkey" as the only way to stop the arrival of the hedonic "false cues" associated with use.

The supposition that consumers choose with foresight is the driving force behind the main empirical prediction of the rational addiction literature: that *future* price increases will generate a decrease in *current* consumption. Several studies have borne out this prediction by measuring, for example, the effect of announced (but not yet effective) increases in cigarette taxes.[35] We acknowledge this empirical phenomenon, but would also suggest an alternative explanation: it could well be that, coincident with the announcement of tax increases on cigarettes, there is an

---

both the role of opioids in adaptive pain management (ter Riet et al., 1998) and their interaction with the dopaminergic system (Deth, 2003).

[35] The most compelling support for this phenomenon is provided by Gruber and Köszegi (2001); see also references therein.

increase in public awareness (due, perhaps, to increased news coverage or increased funding of public health campaigns) of the health consequences of smoking—indeed, such public awareness might well precede and precipitate legislative action. We leave the question of the relative importance of these competing hypotheses (foresight vs. awareness) for future research.

4.3. Implications for public policy

When the time comes to translate a carefully crafted economic theory of addiction into recommendations for public policy, it quickly becomes clear that core assumptions about information and personal responsibility drive everything. After all, the fact that drugs (both legal and illicit) are bought and sold in the marketplace is what motivates the use of economic analysis in the first place, and the primacy of the market mechanism in allocating economic resources is beyond dispute. If it were true that consumers choose to consume addictive substances with complete foresight, without uncertainty or self-control problems, there would seem to be little justification for government to interfere with market transactions.[36] But with uncertainty, incomplete information, and time inconsistency, a role for policy is introduced for "paternalistic" reasons that do not necessarily apply to ordinary consumer goods.[37] This aspect of drug policy finds support within our framework. While it is hard to deny that the lives of heroin junkies might be improved by restrictions on drug availability, our findings suggest that this same line of reasoning could be applied to more mundane objects of consumption such as sugar-coated cereals, lollipops, and wine coolers.

We expect that our findings will come as good news to advocates of public health. We want to be quick to note, however, that we do not mean to imply that individuals who choose to smoke, or drink, or eat sweets are necessarily *wrong*, in the personal sense, to do so. First of all, the discrepancy we noted between the objective function v_t and the subjective function $\tilde{v}_t$ in Section 3.2.1 is to some extent irrelevant to the considerations of human welfare implicit in economic analysis of public policy. Conventional economic analysis rightly focuses on "as if" utility maximization—the equivalent of our subjective function—because in most cases it represents the best available proxy for the real pleasure and pain experienced by consumers in everyday life. While it is evident, given the historical novelty of modern drug technology, that the long-term negative consequences of drug addiction are likely—in the absence of concerted efforts to educate the populace—to be systematically underestimated by the average user,[38] the trade-off between immediate pleasure and future costs is just that: a trade-off. And it is not hard

[36] This is strictly true, of course, only under idealized market conditions. If, for example, the additional healthcare costs incurred by a smoker are covered by insurance, considerations of economic efficiency would dictate that the smoker incur an equivalent cost (in the form of, say, a cigarette tax or increased insurance premiums) contingent on his decision to smoke. The possibility of external effects (e.g., crime, or second-hand smoke) imposed by addicts on others could also provide justification for market intervention. Such considerations are not unique to addictive substances (and therefore will be largely neglected in the present analysis), but they would certainly warrant attention in a more complete analysis of drug policy.

[37] Such policies might include, for example, taxes on drugs (Fehr and Zych, 1998; Gruber and Köszegi, 2001), public education campaigns (Orphanides and Zervos, 1995; Bernheim and Rangel, 2002), restrictions on advertising (Orphanides and Zervos, 1995; Laibson, 2001; Bernheim and Rangel, 2002), regulation of drug dispensation (Bernheim and Rangel, 2002; Camerer et al., 2003), restrictions on public consumption (Laibson, 2001; Bernheim and Rangel, 2002), rehabilitation programs (Orphanides and Zervos, 1995), and criminalization (Bernheim and Rangel, 2002).

[38] It has been noted (by, for instance, Kahneman et al., 1997), that consumer behavior often appears to be inconsistent with the intertemporal maximization of hedonic experience. Biased or false subjective "beliefs" of the sort implied by our subjective function $\tilde{v}_t$ may provide one source of such inconsistency.

to imagine circumstances in which indulging in addiction might, on balance, make an individual better off. Indeed, such circumstances are suggested by changes in the incidence of cigarette smoking in the US in the past four decades: as increasing regulation and aggressive public education campaigns have sharply reduced the prevalence of smoking,[39] the incidence of smoking has become increasingly concentrated in those with lower socioeconomic status and in individuals suffering from such behavioral or affective disorders as depression, adult attention deficit hyperactivity disorder, anxiety disorders, and bulimia. In an insightful review of the medical literature, Pomerleau (1997) has argued that in each of these cases nicotine dependence appears to ameliorate the symptoms, making life more livable for the afflicted.[40] In other words, though smoking might play no beneficial role in a perfect world where health and happiness reign supreme, that is not the world in which we live.

5. Conclusion

That human behavior can usefully be thought of as "rational" is a central tenet of economic theory. But what do we mean by rationality? In the popular lexicon, rationality is coolly deliberate, conscious decision-making. This standard of rationality is clearly not met for many users of illicit drugs. A less rigorous standard is "as if" rationality—satisfied if observed behavior is consistent with the solution to an optimization problem. Is this definition of rationality met for drug addiction? It depends very much on how the problem is specified: if one is willing to be flexible with the domain of preferences, with time inconsistency, and with uncertainty and prior beliefs, then surely *any* pattern of behavior can be justified as the solution to an optimization problem. This is not to say that the rigorous scientific debate over the essential properties of a positive economic theory of addiction has not produced useful insights—on the contrary, this process is the lifeblood of scientific knowledge. Nevertheless, it is our hope that by providing the beginnings of a biological foundation for the theory of rational addiction, we will have helped in some small way to better inform the debate. The framework developed here provides support for the notion that adjacent complementarity can be expected to be important in some consumptive behaviors, and that addiction might indeed be related to problems with self-control, to emotional mechanisms, and to false prior beliefs. But making sense of these disparate behavioral phenomena is easier when we acknowledge directly that what we observe is the manifestation of a sophisticated biological system in which environmental cues trigger predictable internal neurological and physiological responses; that this system shows all the signs of being adapted to a pre-industrial environment; and that drugs of abuse, largely developed in the modern era, have the demonstrable ability to disrupt this system.

It has not escaped our notice that the mismatch we have postulated between human biology and the modern marketplace might extend beyond the realm of drugs and foodstuffs. In this essay, we began with a narrow question: Why is every human being on the planet endowed at birth with an endogenous opioid system, making each and every one of us susceptible to the effects of heroin and other drugs? We then argued that although the evidence suggests this particular component of the human nervous system evolved for a particular purpose (choosing a balanced

---

[39] Between 1965 and 1990, for example, smoking among US adults declined from 40 to 29% (US Department of Health and Human Services, 1989).

[40] Though the "self-medication" role for psychotropic substances remains controversial, few would argue that the pain-relieving function of drugs such as morphine has not proved beneficial in the practice of modern medicine.

diet), this system can be "hijacked" by technological advances such as the syringe, refined sugar, and television advertising. These observations have a very specific implication for economic decision theory (to wit, the persistence of false beliefs), which—though it represents a departure from convention—leads to a uniquely parsimonious explanation of many aspects of both dietary choice and substance abuse. Given that there are other important ways in which pre-industrial environments differed from the modern world, and that there are many other peculiarities of human behavior that provide fodder for the laboratories of behavioral economists, it might be informative to investigate the natural origins of the molecular systems involved. Needless to say, the links between ancestral environments, the human genome, and modern health and well-being will always be indirect and will in every instance require a synthesis of evidence from a broad array of disciplines. But this does not mean the links are not real, and the alternative—ignoring or dismissing such evidence as irrelevant or peripheral—is not likely to yield a sustainable science of economics.

Appendix A

Proof of Proposition 1. If the agent chooses bundle (x, a), then the nutritional content is $C_x x + C_a a$, which has density function

$$
\begin{aligned}
h(t) &= \int_{-\infty}^{\infty} \frac{1}{ax} f_x\left(\frac{y}{x}\right) f_a\left(\frac{t-y}{a}\right) dy \\
&= \int_{\max\{0, t-a\}}^{\min\{x, t\}} \frac{\beta}{ax} \left(\frac{t-y}{a}\right)^{\beta-1} dy \\
&= \left[-\frac{1}{x} \left(\frac{t-y}{a}\right)^{\beta} \right]_{\max\{0, t-a\}}^{\min\{x, t\}}.
\end{aligned}
$$

Carrying out the substitutions we obtain

$$
h(t) = \begin{cases}
\frac{1}{x}(\frac{t}{a})^\beta & \text{if } x \leqslant a, 0 \leqslant t < x, \\
\frac{1}{x}(\frac{t}{a})^\beta - \frac{1}{x}(\frac{t-x}{a})^\beta & \text{if } x \leqslant a, x \leqslant t < a, \\
\frac{1}{x} - \frac{1}{x}(\frac{t-x}{a})^\beta & \text{if } x \leqslant a, a \leqslant t \leqslant x+a, \\
\frac{1}{x}(\frac{t}{a})^\beta & \text{if } a < x, 0 \leqslant t < a, \\
\frac{1}{x} & \text{if } a < x, a \leqslant t < x, \\
\frac{1}{x} - \frac{1}{x}(\frac{t-x}{a})^\beta & \text{if } a < x, x \leqslant t \leqslant x+a, \\
0 & \text{if } t < 0 \text{ or } x+a < t.
\end{cases}
$$

Now in order to derive the agent's objective function we must determine $P(C_x x + C_a a \geqslant k) = \int_k^\infty h(t)\,dt$. In particular,

$$
P(C_x x + C_a a \geqslant k) = \begin{cases}
0 & \text{if } a+x \leqslant k, \\
1 - \frac{k-a}{x} - \frac{1}{\beta+1}\frac{a}{x} + \frac{1}{\beta+1}\frac{a}{x}(\frac{k-x}{a})^{\beta+1} & \text{if } a \leqslant k < a+x, \\
1 - \frac{1}{\beta+1}\frac{a}{x}(\frac{k}{a})^{\beta+1} + \frac{1}{\beta+1}\frac{a}{x}(\frac{k-x}{a})^{\beta+1} & \text{if } x \leqslant k < a, \\
1 - \frac{1}{\beta+1}\frac{a}{x}(\frac{k}{a})^{\beta+1} & \text{if } 0 \leqslant k < x, \\
1 & \text{if } k < 0,
\end{cases}
$$

whenever $x \leqslant a$, and

$$
P(C_x x + C_a a \geqslant k) = \begin{cases}
0 & \text{if } a+x \leqslant k, \\
1 - \frac{k-a}{x} - \frac{1}{\beta+1}\frac{a}{x} + \frac{1}{\beta+1}\frac{a}{x}(\frac{k-x}{a})^{\beta+1} & \text{if } x \leqslant k < a+x, \\
1 - \frac{k-a}{x} - \frac{1}{\beta+1}\frac{a}{x} & \text{if } a \leqslant k < x, \\
1 - \frac{1}{\beta+1}\frac{a}{x}(\frac{k}{a})^{\beta+1} & \text{if } 0 \leqslant k < a, \\
1 & \text{if } k < 0,
\end{cases}
$$

whenever $x > a$. The indifference curves can be divided into five distinct regions, which we illustrate in Fig. 3. The death zone $A^0 = \{(x,a) \in \mathbb{R}_+^2 \mid a + x \leqslant k\}$ in which the probability of survival equals zero, the low-survival region $A^{--} = \{(x,a) \in \mathbb{R}_+^2 \mid k < a+x, a \leqslant k, x \leqslant k\}$ in which survival probability is positive but the consumption levels of both goods are insignificant (i.e., $a, x \leqslant k$), the region $A^{-+} = \{(x,a) \in \mathbb{R}_+^2 \mid k < a, x \leqslant k\}$ in which the consumption level of a is significant while that of x is not, the region $A^{+-} = \{(x,a) \in \mathbb{R}_+^2 \mid a \leqslant k, k < x\}$ in which the consumption level of x is significant while that of a is not, and the region $A^{++} = \{(x,a) \in \mathbb{R}_+^2 \mid k < a, k < x\}$ in which the consumption levels of both x and a are significant.[41] With the exception of area A^0, where the region of indifference consists of the entire area of A^0, the level curves going through regions A^{--}, A^{-+}, A^{+-} and A^{++} are, as can be verified, strictly decreasing and continuously differentiable.

For our purposes, region A^{++} will play the major role, and therefore we describe the indifference curves passing through this area in detail. In particular, within A^{++} we have hyperbolic indifference curves given by

$$
x = \frac{1}{(1-q)(\beta+1)}\frac{k^{\beta+1}}{a^\beta}, \tag{1}
$$

---

[41] We want to emphasize here that the regions depicted in Fig. 3 do not only arise for the specific distribution functions specified by Case 1, but arise in many other cases—for instance, whenever Assumption 1 is satisfied.
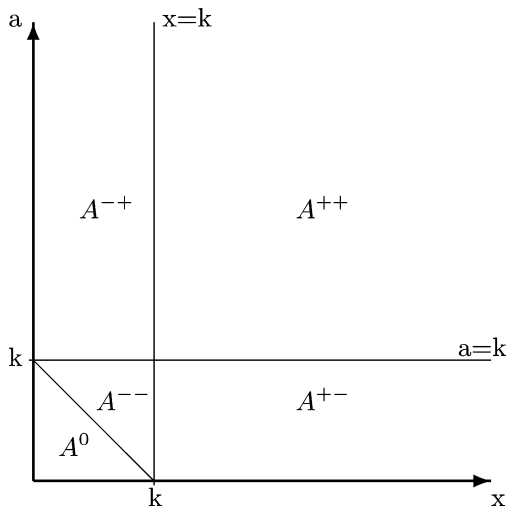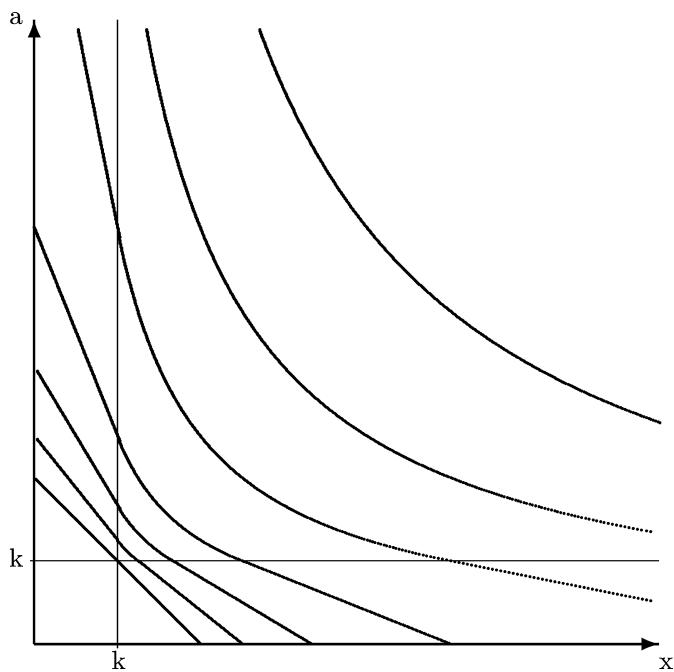
Fig. 3. Five regions.

Fig. 4. Indifference curves ($\beta = 1$, $k = 1$).

where the probability of survival q must lie in $(\frac{\beta}{\beta+1}, 1)$. As can be easily checked, the indifference curves are convex within areas A^{+-} and A^{++}. We show in Fig. 4 the indifference curves associated with survival probabilities 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 and 0.975, with parameter values $\beta = 1$ and $k = 1$.

An important property of the points lying in area A^{++} is that the slope of the indifference curves going through a fixed point $(x, a) \in A^{++}$ increases (or, stated differently, decreases in absolute value) as β increases. This can be verified by first calculating the following derivative in implicit form

$$\frac{da}{dx} = -\frac{\frac{\partial}{\partial x} P(C_x x + C_a a \geqslant k)}{\frac{\partial}{\partial a} P(C_x x + C_a a \geqslant k)} = -\frac{\frac{1}{1+\beta} k^{1+\beta} \frac{1}{x^2 a^\beta}}{\frac{1}{1+\beta} k^{1+\beta} \frac{\beta}{xa^{1+\beta}}} = -\frac{a}{\beta x}$$

and then calculating

$$\frac{\partial}{\partial \beta}\left(-\frac{a}{\beta x}\right) = \frac{a}{x\beta^2} > 0. \tag{2}$$

We can conclude by (2) that a positive cue increases the consumed amount of good a, since as we already know the indifference curves are decreasing, continuous and moreover convex within A^{++}. Thus, a positive cue causes the agent to move in the direction of increasing a along the budget line. The agent might even move into region A^{-+} (indeed, the possibility of a corner solution arises here, as indifference curves in this region can be concave for parameter values $\beta \in (0, 1)$), but this causes no problem because in this case we would have an even greater increase in the consumption of good a. □

From (1) one can see that in A^{++} the indifference curves are Cobb–Douglas indifference curves.[42] Hence, we can formulate the following corollary:

Corollary 2. *Under the assumptions of Proposition* 1*, the objective function* $P(C_x x + C_a a \geqslant k)$ *is observationally equivalent to a Cobb–Douglas utility function within area* A^{++}.

Before we can proceed with the proof of Proposition 2 we need to define the notions of ε-benefit and ε-loss formally. If the agent switches from bundle (x, a) to bundle $(x - \varepsilon, a + \varepsilon/p)$, where $\varepsilon > 0$, he increases the survival area in (c_x, c_a)-space by

$$B := \left\{(c_x, c_a) \in \mathbb{R}_+^2 \mid 0 \leqslant c_x \leqslant \frac{k}{pa+x}, \ \frac{pk}{pa+\varepsilon} - c_x \frac{p(x-\varepsilon)}{pa+\varepsilon} \leqslant c_a \leqslant \frac{k}{a} - c_x \frac{x}{a}\right\},$$

while he decreases the survival area by

$$L := \left\{(c_x, c_a) \in \mathbb{R}_+^2 \mid 0 \leqslant c_a \leqslant \frac{k}{pa+x}, \ \frac{k}{x} - c_a \frac{a}{x} \leqslant c_x \leqslant \frac{k}{x-\varepsilon} - c_a \frac{pa+\varepsilon}{p(x-\varepsilon)}\right\}$$

(see Fig. 2). We refer to the increased probability of survival attributable to the additional area B as ε-*benefit* and to the decreased probability of survival attributable to the lost area L as ε-*loss*.

Remark 1. For any given $\varepsilon \in (0, x)$ we can calculate the ε-benefit by

$$\int_0^{\frac{k}{pa+x}} \int_{\frac{pk}{pa+\varepsilon} - c_x \frac{p(x-\varepsilon)}{pa+\varepsilon}}^{\frac{k}{a} - c_x \frac{x}{a}} f_x(c_x) \hat{f}_a(c_a) \, dc_a \, dc_x$$

---

[42] The indifference curves of a Cobb–Douglas utility function $U(x, a)$ take the form $x^\gamma a^\delta = q$, where $\gamma > 0$ and $\delta > 0$ are parameters and q is a constant corresponding to the level. Any function with level curves that can be represented in this way is observationally equivalent to a Cobb–Douglas utility function.

and the ε-loss by

$$\int_0^{\frac{kp}{pa+\varepsilon}} \int_{\frac{k}{x}-c_a\frac{a}{x}}^{\frac{k}{x-\varepsilon}-c_a\frac{pa+\varepsilon}{p(x-\varepsilon)}} f_x(c_x)\hat{f}_a(c_a)\,\mathrm{d}c_x\,\mathrm{d}c_a.$$

Proof of Proposition 2. The monotonicity of the objective functions $P(C_x x + C_a a \geqslant k)$ and $P(C_x x + \widehat{C}_a a \geqslant k)$ in (x, a), outside the area A^0, imply that the indifference curves associated with a positive probability of survival are strictly decreasing curves and that the agent will select in both cases a bundle lying on his budget line $x + pa = m$.

Since $P(C_x x + C_a a \geqslant k) = \int_k^{x+a} \int_{\max\{0,t-a\}}^{\min\{x,t\}} \frac{1}{ax} f_x(\frac{y}{x}) f_a(\frac{t-y}{a})\,\mathrm{d}y\,\mathrm{d}t$ is a continuous function in (x, a), the upper contour sets of the objective function are closed. This, and knowing that the indifference curves associated with the objective function $P(C_x x + C_a a \geqslant k)$ are strictly decreasing curves in case of positive survival probabilities, implies that these indifference curves must be continuous. For the same reasons $P(C_x x + \widehat{C}_a a \geqslant k)$ must also have continuous indifference curves.

If there exists an $\varepsilon \in (0, x^*)$ such that the ε-benefit exceeds the ε-loss, then

$$P\big(C_x x^* + \widehat{C}_a a^* \geqslant k\big) = P(L \cup C) < P(B \cup C)$$

$$= P\left(C_x(x^* - \varepsilon) + \widehat{C}_a\left(a^* + \frac{\varepsilon}{p}\right) \geqslant k\right), \tag{3}$$

where

$$C = \left\{(c_x, c_a) \in \mathbb{R}_+^2 \mid c_x x^* + c_a a^* \geqslant k \text{ and } c_x(x^* - \varepsilon) + c_a\left(a^* + \frac{\varepsilon}{p}\right) \geqslant k\right\},$$

and therefore, the agent increases his probability of survival by exchanging $(x^* - \varepsilon, a^* + \varepsilon/p)$ for (x^*, a^*). The single-crossing property imposed on the two indifference curves associated with the no-cue balanced diet problem and the positive-cue balanced diet problem that pass through (x^*, a^*), together with Eq. (3) implies that the former indifference curve must lie below the latter whenever $x > x^*$. Hence, consuming less than a^* in the positive-cue balanced diet problem will be dominated by the allocation $a^* + \varepsilon/p$, and the solution to the positive-cue balanced diet problem necessarily occurs where $\hat{a} > a^*$.

It remains to be shown that if there does *not* exist an $\varepsilon \in (0, x^*)$ such that the ε-benefit exceeds the ε-loss, then the solution to the positive-cue balanced diet problem will occur where $\hat{a} \leqslant a^*$. Noting that this implies that expression (3) cannot be satisfied, $\hat{a} \leqslant a^*$ is implied by the uniqueness of the solution $\hat{a}$. □

Proof of Proposition 3. First, we verify that a positive cue increases the expected probability of the occurrence of the cue, i.e., $E\Pi_t < E\Pi_{t+1}$. To see this multiply both sides of Eq. (3) by π and thereafter integrate both sides with respect to π to get

$$E\Pi_{t+1} = \int_0^1 \pi g_{t+1}(\pi) = \frac{\int_0^1 \pi^2 g_t(\pi)\,\mathrm{d}\pi}{\int_0^1 \pi g_t(\pi)\,\mathrm{d}\pi} = \frac{E\Pi_t^2}{E\Pi_t}. \tag{4}$$

Now, employing a Jensen-type inequality for strictly convex functions and for non-degenerate as well as non-negative random variables we obtain $E\Pi_t < E\Pi_{t+1}$ by (4).

For any point $(m - pa, a)$ lying on the budget line we obtain from the definitions of v_t and v_{t+1} (suppressing the arguments $(m - pa, a)$) that

$$\frac{\mathrm{d}}{\mathrm{d}a} v_{t+1} - \frac{\mathrm{d}}{\mathrm{d}a} v_t = (E\Pi_{t+1} - E\Pi_t)\left(\frac{\partial u}{\partial a} - p\frac{\partial u}{\partial x} - \left(\frac{\partial \bar{u}}{\partial a} - p\frac{\partial \bar{u}}{\partial x}\right)\right). \tag{5}$$

Noting that the first factor of the right-hand side of (5) is positive, we now need to show that the second factor is positive. This is equivalent to demonstrating the inequality

$$\left(\frac{\partial u/\partial a}{\partial u/\partial x} - p\right)\frac{\partial u}{\partial x} > \left(\frac{\partial \bar{u}/\partial a}{\partial \bar{u}/\partial x} - p\right)\frac{\partial \bar{u}}{\partial x} \tag{6}$$

at $(m - pa, a)$. We have $\frac{\partial u}{\partial x}(m - pa, a) > 0$ and $\frac{\partial \bar{u}}{\partial x}(m - pa, a) > 0$ by the monotonicity of the utility functions outside the death zone A^0.

Since we assumed that the optimal solutions associated with the no-cue balanced diet problem and the positive-cue balanced diet problem have unique solutions in A^{++}, we get $(x^{\bar{c}}, a^{\bar{c}})$ and (x^c, a^c) by solving the respective first-order conditions. Hence,

$$\frac{\partial u/\partial a(x^c, a^c)}{\partial u/\partial x(x^c, a^c)} = p \quad \text{and} \quad \frac{\partial \bar{u}/\partial a(x^{\bar{c}}, a^{\bar{c}})}{\partial \bar{u}/\partial x(x^{\bar{c}}, a^{\bar{c}})} = p, \tag{7}$$

where $x^c = m - pa^c$ and $x^{\bar{c}} = m - pa^{\bar{c}}$ by the monotonicity of the objective functions. Therefore, as illustrated in Fig. 5, it follows from strict concavity of the objective functions that $\frac{\partial u}{\partial a}/\frac{\partial u}{\partial x} > p$ if $k < a < a^c$, $\frac{\partial u}{\partial a}/\frac{\partial u}{\partial x} < p$ if $a > a^c$, $\frac{\partial \bar{u}}{\partial a}/\frac{\partial \bar{u}}{\partial x} > p$ if $k < a < a^{\bar{c}}$ and $\frac{\partial \bar{u}}{\partial a}/\frac{\partial \bar{u}}{\partial x} < p$ if $a > a^{\bar{c}}$ at a point $(m - pa, a)$. Thus, (6) is satisfied for all $a \in [a^{\bar{c}}, a^c]$.

Finally, we need to verify our statement on the solution (x^*, a^*) of problem (2). Observe that the optimal solution (x^*, a^*) at time t must lie on the budget line by the monotonicity of the objective function of (2). Hence, $x^* = m - pa^*$. Since u and $\bar{u}$ are strictly concave in A^{++}, it
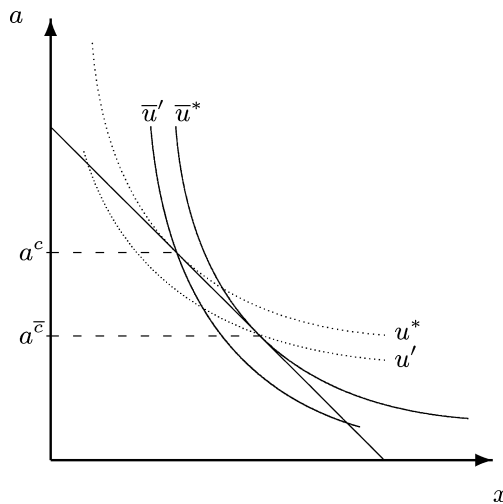

Fig. 5. Intermediate optimum.

follows that v_t is also strictly concave in A^{++}. Therefore, and by the assumption that the optimal solution associated with problem (2) lies also in A^{++}, the first-order condition

$$E\Pi_t\left(\frac{\partial u/\partial a}{\partial u/\partial x} - p\right)\frac{\partial u}{\partial x} + (1 - E\Pi_t)\left(\frac{\partial \bar{u}/\partial a}{\partial \bar{u}/\partial x} - p\right)\frac{\partial \bar{u}}{\partial x} = 0$$

of problem (2) determines $(x^*, a^*) = (m - pa^*, a^*)$ and we see that

$$\frac{\partial u/\partial a}{\partial u/\partial x} - p \quad \text{and} \quad \frac{\partial \bar{u}/\partial a}{\partial \bar{u}/\partial x} - p \tag{8}$$

cannot have the same sign.[43] However, this implies—as illustrated in Fig. 5—that $a^* \in (a^{\bar{c}}, a^c)$, because otherwise the two expressions in (8) would have identical signs. $\quad\square$

Proof of Proposition 4. Suppose that the agent maximizes his subjective function $\tilde{v}_t$ instead of v_t in problem (2). Observe that for the subjective function $\tilde{v}_t$ we can apply Proposition 3. Hence, subjective adjacent complementarity is satisfied at $(\tilde{x}_t, \tilde{a}_t)$ and we must have $0 = \frac{d}{da}\tilde{v}_t(m - p\tilde{a}_t, \tilde{a}_t) < \frac{d}{da}\tilde{v}_{t+1}(m - p\tilde{a}_t, \tilde{a}_t)$. Now since $\tilde{a}_{t+1}$ is the unique solution of $\frac{d}{da}\tilde{v}_{t+1}(m - pa, a) = 0$ within the region $x, a > k$, it follows that $\tilde{a}_t < \tilde{a}_{t+1}$. It remains to be shown that the agent's true expected survival probability given by $v_t(m - p\tilde{a}_t, \tilde{a}_t)$ decreases strictly in time. Clearly, the agent is moving along the budget line farther and farther away from his true expected survival maximizing bundle $(x^{\bar{c}}, a^{\bar{c}})$ because $\tilde{a}_t$ increases. This implies by the strict concavity of $v_t = \bar{u}$ that $v_t(\tilde{x}_t, \tilde{a}_t) > v_{t+1}(\tilde{x}_{t+1}, \tilde{a}_{t+1})$. $\quad\square$

References

Ainslie, G., 1991. Derivation of rational economic behavior from hyperbolic discount curves. Amer. Econ. Rev. 81 (2), 334–340.

Atkinson, R., 1982. Naloxone decreases food intake in obese humans. J. Clinical Endocrinology Metabolism 55, 196–198.

Baldwin, B., Parrot, R., 1985. Effects of intracerebroventricular injection of naloxone on operant feeding and drinking in pigs. Pharmacol. Biochem. Behav. 22, 37–40.

Baron-Cohen, S., 1995. Mindblindness: An Essay on Autism and Theory of Mind. MIT Press, Cambridge, MA.

Becker, G.S., Murphy, K.M., 1988. A theory of rational addiction. J. Polit. Economy 96 (4), 675–700.

Becker, G.S., Grossman, M., Murphy, K.M., 1994. An empirical analysis of cigarette addiction. Amer. Econ. Rev. 84, 396–418.

Bergstrom, T., 1996. Economics in a family way. J. Econ. Lit. 34 (4), 1903–1934.

Bernheim, B.D., Rangel, A., 2002. Addiction and cue conditioned cognitive processes. Working paper w9329. NBER.

Blass, E., 1986. Functional interaction between the positive effect of sweet and the negative effect of pain and distress. Appetite 7, 243.

Blass, E., Hoffmeyer, L., 1991. Sucrose as an analgesic for newborn infants. Pediatrics 87, 215–218.

Blass, E., Fitzgerald, E., Kehoe, P., 1987. Interactions between sucrose, pain and isolation distress. Pharmacol. Biochem. Behav. 26, 483–489.

Bodnar, R.J., Hadjimarkou, M.M., 2002. Endogenous opiates and behavior: 2001. Peptides 23, 2307–2365.

Booth, M., 1996. Opium: A History. Simon & Schuster, London.

Brick, J., Erickson, C., 1998. Drugs, the Brain, and Behavior: The Pharmacology of Abuse and Dependence. Haworth Medical Press, Binghamton, New York.

Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T., Rabin, M., 2003. Regulation for conservatives: Behavioral economics and the case for asymmetric paternalism. Univ. Pennsylvania Law Rev. 151, 1211–1254.

Camerer, C., Loewenstein, G., Prelec, D., 2005. Neuroeconomics: How neuroscience can inform economics. J. Econ. Lit. 43 (1), 9–64.

---

[43] Observe that the two terms in (8) cannot equal zero at the same point on the budget line because of $a^{\bar{c}} < a^c$ and (7).

Campbell, M., Adams, P., Small, A., Yesch, L., Curren, E., 1988. Naltrexone in infantile autism. Psychopharmacol. Bull. 24, 135–139.

Carson-DeWitt, R., 2001. Encyclopedia of Drugs, Alcohol, and Addictive Behavior, second ed. Macmillan Reference USA, New York.

Chiara, G.D., 1999. Drug addiction as a dopamine-dependent associative learning disorder. Europ. J. Pharmacol. 375, 13–30.

Cohen, M., Cohen, R., Pickar, D., Murphy, L., 1985. Naloxone reduces food intake in humans. Psychosomatic Medicine 47, 132–138.

Colantuoni, C., Rada, P., McCarthy, J., Patten, C., Avena, N.M., Chadeayne, A., Hoebel, B.G., 2002. Evidence that intermittent, excessive sugar intake causes endogenous opioid dependence. Obesity Res. 10, 478–488.

Cooper, S., Higgs, S., 1994. Neuropharmacology of appetite and taste preferences. In: Legg, C., Booth, D. (Eds.), Appetite, Neural and Behavioral Bases. Oxford Univ. Press, New York, pp. 212–242.

Cooper, J.R., Bloom, F.E., Roth, R.H., 2002. The Biochemical Basis of Neuropharmacology, eighth ed. Oxford Univ. Press.

Cummings, D.E., Schwartz, M.W., 2003. Genetics and pathophysiology of human obesity. Annual Rev. Medicine 54, 453–471.

Deth, R.C., 2003. Molecular Origins of Human Attention: The Dopamine-Folate Connection. Kluwer Academic, Boston.

Drewnowski, A., Krahn, D., Demitrack, M., Nairn, K., Gosnell, B., 1992. Taste responses and preferences for sweet high-fat foods: Evidence for opioid involvement. Physiol. Behav. 51, 371–379.

Dudley, R., 2000. Evolutionary origins of human alcoholism in primate frugivory. Quart. Rev. Biol. 75 (1), 315.

Dum, J., Gramsch, C., Herz, A., 1983. Activation of hypothalamic beta-endorphin pools by reward induced by highly palatable food. Pharmacol. Biochem. Behav. 18, 443–447.

Fantino, M., Hosotte, J., Apfelbaum, M., 1986. An opioid antagonist, naltrexone, reduces preference for sucrose in humans. Amer. J. Physiol. 252, R91–R96.

Fehr, E., Zych, P.K., 1998. Do addicts behave rationally? Scand. J. Econ. 100 (3), 643–662.

Foster, J., Morrison, M., Dean, S., Hill, M., Frenk, H., 1981. Naloxone suppresses food/water consumption in the deprived cat. Pharmacol. Biochem. Behav. 14, 419–421.

Gillberg, C.L.T., Lonnerholm, G., 1985. Endorphin activity in childhood psychosis. Archives General Psychiatry 42, 780–783.

Gilbertson, T., Kim, I., 2002. 'Chemoreception of fat,' chemistry of taste: Mechanisms, behaviors, and mimics. ACS Symp. Ser. 825, 180–191.

Gosnell, B., Levine, A., 1996. Stimulation of ingestive behavior by preferential and selective opioid agonists. In: Cooper, S., Clifton, P. (Eds.), Drug Receptor Subtypes and Ingestive Behaviour. Academic Press, San Diego, pp. 147–166.

Griffiths, R., Balster, R., 1979. Opioids: Similarity between evaluations of subjective effects and animal self-administration results. Clinical Pharmacol. Therapeutics 25, 611–617.

Grossman, M., Chaloupka, F.J., Sirtalan, I., 1998. An empirical analysis of alcohol addiction: Results from monitoring the future panels. Econ. Inquiry 36 (1), 39–48.

Gruber, J., Köszegi, B., 2001. Is addiction rational? Theory and evidence. Quart. J. Econ. 116 (4), 1261–1303.

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

Harrigan, S., Downs, D., 1978. Continuous intravenous naltrexone effects on morphine self-administration in rhesus monkeys. J. Pharmacol. Exper. Therapeutics 204, 481–486.

Headlee, C., Coppock, H., Nichols, J., 1955. Apparatus and technique involved in a laboratory method of detecting the addictiveness of drugs. J. Amer. Pharmaceutical Assoc. 44, 229–231.

Hirshleifer, J., 1977. Economics from a biological viewpoint. J. Law Econ. 20, 1–52.

Hirshleifer, J., 1985. The expanding domain of economics. Amer. Econ. Rev. 75 (6), 53–68.

Holtzman, S., 1974. Behavioral effects of separate and combined administration of naloxone and d-amphetamine. J. Pharmacol. Exper. Therapy 189, 51–60.

Jaffe, J., Martin, W., 1990. Opioid analgesics and antagonists. In: Gillmina, A., Rall, T., Nies, A., Taylor, P. (Eds.), The Pharmacological Basis of Therapeutics. Pergamon Press, New York, pp. 485–521.

Kahneman, D., Wakker, P.P., Sarin, R., 1997. Back to Bentham? Explorations of experienced utility. Quart. J. Econ. 112 (2), 375–405.

Kalat, J.W., 1978. Letter to the editor: Speculations on similarities between autism and opiate addiction. J. Autism Childhood Schizophrenia 8 (4), 477–479.

Kavaliers, M., Hirst, M., Teskey, G., 1985. Opioid systems and feeding in the slug: Similarities to and implications for mammalian feeding. Brain Res. Bull. 14, 681–685.

Kavaliers, M., Guglick, M., Hirst, M., 1987. Opioid involvement in the control of feeding of an insect, the American cockroach. Life Sci. 49, 665–672.

Killian, A., Bonese, K., Schuster, C., 1978. The effects of naloxone on behavior maintained by cocaine and heroin injections in the rhesus monkey. Drug Alcohol Depend. 3, 243–251.

Laibson, D., 1997. Golden eggs and hyperbolic discounting. Quart. J. Econ. 112 (2), 443–477.

Laibson, D., 2001. A cue-theory of consumption. Quart. J. Econ. 116 (1), 81–119.

Leboyer, M., Bouvard, M., Dugas, M., 1988. Effects of naltrexone on infantile autism. Lancet 1, 715.

Leboyer, M., Bouvard, M., Lensing, P., Launay, J., 1990. Opioid excess hypothesis of autism: A double-blind study of naltrexone. Brain Dysfunction 3, 285–298.

Loewenstein, G., 1996. Out of control: Visceral influences on behavior. Organ. Behav. Human Dec. Processes 65 (3), 272–292.

Martin, W., Wickler, C., Eades, C., Pescor, F., 1963. Tolerance to and physical dependence on morphine in rats. Psychopharmacologia 4, 247–260.

McKim, W.A., 2002. Drugs and Behavior: An Introduction to Behavioral Pharmacology, fifth ed. Prentice Hall.

Mello, N.K., Negus, S.S., 1996. Preclinical evaluation of pharmacotherapies for treatment of cocaine and opioid abuse using drug self-administration procedures. Neuropsychopharmacology 14 (6), 357–424.

Mercer, M.E., Holder, M.D., 1997. Food cravings, endogenous opioid peptides, and food intake: A review. Appetite 29, 325–352.

Morley, J., Parker, S., Levine, A., 1985. The effect of butorphanol tartrate on food and water consumption in humans. Amer. J. Clinical Nutrition 42, 1175–1178.

Mucha, R., Iversen, S., 1984. Reinforcing properties of morphine and naloxone revealed by conditioned place preferences: A procedural examination. Psychopharmacology 82, 241–247.

Mucha, R.F., van der Kooy, D., O'Shaughnessy, M., Bucenieks, P., 1982. Drug reinforcement studied by the use of place conditioning in rat. Brain Res. 243, 91–105.

National Institute on Drug Abuse, 2000. Heroin: Abuse and addiction, Research report series, NIDA.

Nelson, R.J., 2000. An Introduction to Behavioral Endocrinology, second ed. Sinauer, Sunderland, MA.

Nesse, R., 1994. An evolutionary perspective on substance abuse. Ethol. Sociobiol. 15, 339–348.

Nesse, R.M., Berridge, K.C., 1997. Psychoactive drug use in evolutionary perspective. Science 278, 63–66.

O'Donoghue, T., Rabin, M., 2002. Addiction and present-biased preferences. Mimeo.

Orphanides, A., Zervos, D., 1995. Rational addiction with learning and regret. J. Polit. Economy 103 (4), 739–758.

Orphanides, A., Zervos, D., 1998. Myopia and addictive behaviour. Econ. J. 108, 75–91.

Ostrowski, N., Rowland, N., Foley, T., Nelson, T., Reid, L., 1981. Morphine antagonists and consummatory behavior. Pharmacol. Biochem. Behav. 14, 549–559.

Parfitt, K., 1999. Martindale: The Complete Drug Reference, thirty-second ed. Pharmaceutical Press, Taunton, MA.

Pert, C.B., Snyder, S.H., 1973. Opiate receptor: Demonstration in nervous tissue. Science 179, 1011–1014.

Pomerleau, C.S., 1997. Co-factors for smoking and evolutionary psychobiology. Addiction 92 (4), 397–408.

Raiten, D.J., Massaro, T., 1986. Perspectives on the nutritional ecology of autistic children. J. Autism Developmental Disorders 16 (2), 133–142.

Raven, P.H., Evert, R.F., Eichhorn, S.E., 1999. Biology of Plants, sixth ed. Freeman, New York.

Robson, A.J., 2001. The biological basis of economic behavior. J. Econ. Lit. 39 (1), 11–33.

Rogers, A.R., 1994. Evolution of time preference by natural selection. Amer. Econ. Rev. 84 (3), 460–481.

Rudski, J., Grace, M., Billington, C., Levine, A., 1992. Increased food intake following morphine or MPT administration into the ventral tegmental area of rats. Soc. Neurosci. Abstracts 18, 1427.

Samuelson, L., 2004. Information-based relative consumption effects. Econometrica 72, 93–118.

Sanger, D., McCarthy, P., Lord, J., Smith, C., 1983. The anorectic properties of opiate antagonists. Drugs Devel. Res. 3, 137–142.

Savage, L.J., 1954. The Foundations of Statistics. Wiley.

Schuckit, M.A., 1999. New findings in the genetics of alcoholism. J. Amer. Medical Assoc. 281 (20), 1875–1876.

Si, E., Bryant, H., Yim, G., 1986. Opioid and non-opioid components of insulin-induced feeding. Pharmacol. Biochem. Behav. 24, 899–903.

Simon, E., Hiller, J., Edelman, I., 1973. Stereospecific binding of the potent narcotic analgesic [^{3}H]-etorphine to rat brain homogenate. Proc. Natl. Acad. Sci. 70, 1947–1949.

Simpkins, J., Michel, M., Tuttle, R., 1985. Effects of the narcotic antagonist, nalmefene, on spontaneous and insulin-induced food intake and body weight gain in male rats. Soc. Neurosci. Abstr. 11, 59.

Smith, T.G., 2002. Toward a unified theory of choice: Case studies in dietary preference. Doctoral thesis. University of California, Santa Barbara.

Smith, T.G., 2004. The McDonald's equilibrium: Advertising, empty calories, and the endogenous determination of dietary preferences. Soc. Choice Welfare 23 (3), 383–413.

Sozou, P., 1998. On hyperbolic discounting and uncertain hazard rates. Proc. R. Soc. Lond. Ser. B: Biol. Sci. 265 (1409), 2015–2020.

Spanagel, R., Weiss, F., 1999. The dopamine hypothesis of reward: Past and current status. Trends Neurosci. 22 (11), 521–527.

Sullivan, R., Hagen, E., 2002. Psychotropic substance-seeking: Evolutionary pathology or adaptation? Addiction 97, 389–400.

Sullivan, S.L., 2002. Mammalian chemosensory receptors. NeuroReport 13 (1), A9–A17.

ter Riet, G., de Craen, A.J., de Boer, A., Kessels, A.G., 1998. Is placebo analgesia mediated by endogenous opioids? A systematic review. Pain 76, 273–275.

Terenius, L., 1974. Stereospecific interaction between narcotic analgesics and a synaptic plasma membrane fraction of rat cerebral cortex. Acta Pharmacologica et Toxicologica 32, 317–319.

US Department of Health and Human Services, 1989. Reducing the health consequences of smoking: 25 years of progress. A report of the surgeon general, Publication No. 89-8411, DHHS, CDC.

van Ree, J.M., Wied, D.D., Bradbury, A., Hume, E., Smyth, D., Snell, C., 1976. Induction of tolerance to the analgesic action of lipotropin C-fragment. Nature 264, 792–794.

van Ree, J.M., Gerrits, M.A., Vanderschuren, L.J., 1999. Opioids, reward, and addiction: An encounter of biology, psychology, and medicine. Pharmacol. Rev. 51 (2), 341–396.

van Ree, J.M., Niesink, R.J., van Wolfswinkel, L., Ramsey, N.F., Kornet, M.L.M., van Furth, W.R., Vanderschuren, L.J., Gerrits, M.A., van den Berg, C.L., 2000. Endogenous opioids and reward. Europ. J. Pharmacol. 405, 89–101.

Waeltl, P., Dickinson, A., Schultz, W., 2001. Dopamine responses comply with basic assumptions of formal learning theory. Nature 412, 43–48.

Weeks, J., 1962. Experimental morphine addiction: Method for automatic intravenous injection in unrestrained rats. Science 138, 143–144.

Weeks, J., Collins, R., 1976. Changes in morphine self-administration in rats induced by prostaglandin E1 and naloxone. Prostaglandins 12, 11–19.

Weiss, G., 1982. Food fantasies of incarcerated drug users. Int. J. Addictions 17, 905–912.

Williams, G.C., Nesse, R.M., 1991. The dawn of Darwinian medicine. Quart. Rev. Biol. 66 (1), 1–22.

Yamamoto, T., Sako, N., Maeda, S., 2000. Effects of taste stimulation on beta-endorphin levels in rat cerebrospinal fluid and plasma. Physiol. Behav. 69, 345–350.

Yeomans, M.R., Gray, R.W., 1996. Selective effects of naltrexone on food pleasantness and intake. Physiol. Behav. 60 (2), 439–446.

Yeomans, M.R., Gray, R.W., 2002. Opioid peptides and the control of human ingestive behavior. Neurosci. Biobehav. Rev. 26, 713–728.

Yeomans, M.R., Wright, P., 1991. Lower pleasantness of palatable foods in nalmefene-treated human volunteers. Appetite 16, 249–259.

Zigmond, M.J., Bloom, F.E., Landis, S.C., Roberts, J.L., Squire, L.R., 1999. Fundamental Neuroscience. Academic Press, San Diego.