Original Paper

Bounded Rationality in Compulsive Consumption

Ingo Fiedler¹,²

¹ University of Hamburg, Faculty of Business, Economics & Social Sciences, Hamburg, Germany
² Concordia University, Faculty of Arts & Science, 1455 de Maisonneuve West, Montreal, QC, H3G 1M8, Canada

Received: November 24, 2020     Accepted: December 8, 2020     Online Published: January 30, 2021
doi:10.22158/jbtp.v9n1p30     URL: http://dx.doi.org/10.22158/jbtp.v9n1p30

Abstract

The standard microeconomic assumption is that consumers’ choices maximize consumers’ utility. This theoretical article challenges this assumption by presenting a framework of an extreme case: compulsive consumption. Backed by a wide range of existing empirical evidence it is shown by the example of pathological gamblers that some consumers (1) have inconsistent preferences, (2) underestimate the time horizon and the frequency of consumption, and (3) underestimate the costs and overestimate the benefits of consumption. The results do not necessarily violate rational choice theory if interpreted as intra-personal externalities. By applying the perspective of picoeconomics, it is possible to reconcile fully rational but competing agents within an individual with inconsistent aggregate decisions. Yet, from a welfare perspective, the results imply that bounded rationality can be interpreted as a source of inefficiency and can thus constitute a rationale for regulatory intervention.

Keywords

bounded rationality, addiction, consumer, gambling, picoeconomics, intra-personal externalities

1. Introduction

The assumption of rationality plays a major role in economics and consumer research. It has, however, come under siege from behavioral economists and psychologists, who have shown that humans do not always act fully rationally in every area of life. The policy implications of the two alternative assumptions regarding rationality could not lie further apart: If consumers acted fully rationally, their choices would yield an optimum and any regulation (beyond reducing informational asymmetries and reducing or internalizing externalities) reduces efficiency and overall welfare. If, by contrast, consumers act irrationally from time to time or to a certain degree, then policy intervention can increase welfare, for example by nudging individuals to choose what lies in their best interest. This article presents a theoretical framework and then surveys existing evidence for an extreme case: pathological gambling. It is shown that gamblers’ decisions, especially those of addicted ones, are prone to bounded rationality.

Rationality is defined as consistent utility maximizing behavior. That is, agents who do not act consistently are considered irrational or not fully rational. To structure the theoretical and empirical evidence of bounded rationality, a model is presented that incorporates all costs and benefits of a consumption decision of a good coming with the risk of addiction. Given that in a state of addiction, an individual’s costs of consumption may outweigh the benefits, the model allows to derive four cases of
rational consumption: (1) consumers who are resistant to addiction, (2) only marginally addicted consumers, (3) consumers with a short time horizon, and (4) consumers who made an initial non-systematic error of judgment.

Irrational consumption, on the other hand, can be derived from the model when the players’ (long-term) preference condition systematically fails to coincide with their (short-term) participation condition. Three main sources of such irrational decisions are highlighted: (1) time preferences: steep discounting and/or changing discount functions, (2) underestimation of the time horizon or frequency of consumption, and (3) overestimation of benefits and underestimation of costs. Subsequently, the article presents theoretical and empirical evidence pertaining to these three sources of irrational consumption decisions for the case of gambling. It then suggests that rationality is a continuum of the internalization of private costs and benefits.

The results are discussed in the light of picoeconomics, yielding the idea that irrational behavior on the level of an individual need not violate the overall principle of rationality if behavior is determined by multiple inner agents who are perfectly rational but have competing goals. This interpretation leads to a reconciliation of Rational Choice and Behavioral Economics: While rationality is still valid on the (theoretical) level of the inner agents, there is argumentative scope for real-world policy interventions that address the individual. These findings are discussed with a special focus on social cost-benefit analyses and social welfare, including policy implications like a tax rate in proportion to the addictiveness of a good. The final chapter concludes and presents an outlook for further research.

2. A Theoretical Model of (Ir-)Rational Addictive Consumption

2.1 Modeling Rational Addiction

In 1988, Becker and Murphy proposed the first comprehensive theoretical framework to describe addiction as rational behavior. The authors maintain that “addictions, even strong ones, are usually rational in the sense of involving forward-looking maximization with stable preferences” (Becker & Murphy, 1988, p. 675). Their concept of addictions is wide, encompassing not only substance-based addictions but also behavioral ones like pathological gambling.

Becker and Murphy’s central premise is that the individual’s preference structure is stable over time. The individual’s utility function depends on the consumption of two goods, c and y. c is the addictive good (or behavior); its marginal utility positively depends on its consumption in the past (learning effects). This means that the utility derived from the pleasure of consumption increases with each round of play. The marginal utility of the alternative good y, by contrast, is constant. Total utility is:

\[ U(t) = U[y(t), c(t), S(t)] \]

S is the consumption capital, which accumulates with consumption due to learning effects and changes over time according to the following expression:

\[ \dot{S}(t) = c(t) - \delta S(t) - h[d(t)] \]

\( \delta \) is the exogenous rate of depreciation of the physical and mental effects of past consumption of c, and \( h[d(t)] \) is a function of endogenous expenditure on a reduction of S. If \( \sigma \) defines constant(!) time preferences, then the individual’s utility function, at \( t=0 \) until their end of life at \( T \), is:

\[ U(0) = \int_{0}^{T} e^{-\sigma t} u[y(t), c(t), S(t)] dt \]

Maximizing expression \( (3) \) with consideration to expression \( (2) \) and subject to a budget constraint yields the full price of c, which consists of the sum of its market price and the present value of the future costs of consumption, including the costs of a potential addiction, which must be weighed with a
probability of occurrence.

The model can also be used to show that—keeping future consumption constant—consumption of a harmful good depends positively on $\delta$ and $\sigma$. If the increase in the marginal utility of $c$ over time exceeds the increase in its full price, then adjacent complementarity exists (Ryder & Heal, 1973), the mathematical representation of learning effects, which is an important key to modeling addiction. Becker and Murphy (1988, p. 681) define a person as potentially addicted “if an increase in his current consumption [...] increases his future consumption”. They show that in the framework of their model, adjacent complementarity is a necessary but not a sufficient condition for addiction.

This definition of addiction necessarily implies an interaction between addictive goods and their consumers. Addiction is therefore an individual phenomenon, as evidenced in particular by the role of time preference. Present-oriented individuals are thus at greater risk of addiction than their more future-oriented peers, seeing that the full price of a good increases less steeply if the future is discounted more heavily. Individuality furthermore plays a role in the complementarity between present and past consumption. For example, the effects of consumption on a person’s income depend on both the person and the good.

The principle of adjacent complementarity also predicts that small changes in price will cause large shifts in demand. Becker and Murphy furthermore deduct that changes in the future price of the addictive good will affect its current demand, and in fact even more strongly so than a current price change. According to the authors, this property of addictive goods serves to distinguish rational addiction from myopic behavior, since in the latter, rising future prices would not affect current consumption.

In this article the model of rational consumption of addictive goods is challenged by presenting the case of gambling. Pathological gambling has been classified as a disorder since the third edition of the Diagnostic and Statistical Manual of Mental Disorders of 1980. Like most addictions, pathological gambling entails significant negative consequences for the addict, and virtually all pathological gamblers wish they could escape the addiction or had not begun to gamble in the first place (Petry, 2005; Meyer & Bachmann, 2005). The much higher risk of suicide among pathological gamblers is perhaps the starkest manifestation of their severely reduced quality of life. A model of gambling decisions is developed in the next section and then examined whether gambler’ actual decisions can be fully explained by the concept of rationality or whether they violate one or more of its assumptions.

2.2 Modeling the Decision to Gamble

The utility from gambling, $U_s$, can be attributed to three sources: (1) the monetary value of a win, (2) the pleasure of winning, $\gamma$, and (3) the pleasure of gambling, $\alpha$. The first component depends on the outcome of the game ($W$ for “win” is equal to 1 or 0). The size of the win, measured in multiples, $\lambda$, of the bet, $B$, must therefore be multiplied by the probability of winning, $P_{W=1}$, as must $\gamma$. By contrast, $\alpha$ is independent of the outcome of the game.

The marginal utility of consumption increases along with the consumption capital, $C_i$. This means that $\alpha$ and $\gamma$ positively depend on $C_i$ and, thus, the temporal dimension in the form of $C_i$ and the discount function also enters the utility function:

$$U_{s(t)} = D_i(t) \sum_t^{\infty} \alpha_i C_i(t) + D_i(t) P_{W=1} \sum_t^{\infty} (\gamma_i C_i(t) + \lambda B_i)$$

According to McClure et al. (2007), the discount function, $D_i(t)$, takes the form

$$D_i(t) = \omega(t) \beta^t + (1 - \omega(t)) \delta^t,$$

where $\beta$ is the discount rate for short-term utility and $\delta$ is applied to long-term utility, with $0 \leq \delta < \beta \leq 1$. $\omega(t)$ constitutes the equilibrium function over time between the two discount rates, with $0 \leq \omega \leq 1$. The
further a reward is away, the smaller is \( \omega \). Since \( \beta > \delta \), the “bonus” attributed to immediate rewards decreases over time.

The cost function of gambling decisions, on the other hand, comprises the sum of the bets, \( B_t \), and the costs of addiction, \( k_t \), which assume a wide range of forms (Fiedler, 2016). Moreover, pathological gambling is a continuum (Skog, 2005; Currie & Casey, 2007), and therefore so are the private costs of gambling addiction. It appears reasonable to model these costs as a function of consumption capital (with a positive relationship). Thus, the cost function of gambling is:

\[
K_{\omega(t)} = D_t(t) \left( \sum_t B_t + k_t C_t(t) \right)
\]

Let \( \Psi \) be the house edge. If a decision to gamble is rational, the following inequality holds:

\[
D_t(t)[\sum_t a_t C_t(t) + \sum_{\gamma} P_{\gamma} C_t(t)] > D_t(t) [\sum_t \Psi t + k_t C_t(t)]
\]

If an individual decides to gamble, he increases his consumption capital. The size of the consumption capital depends positively and additively on past consumption, but negatively and multiplicatively on the depreciation rate of the consumption capital. In consequence, the consumption capital does not grow indefinitely but instead approaches an equilibrium level, \( C^* \), at which additions from consumption are exactly offset by depreciation.

Although both the marginal benefit and marginal costs increase in consumption capital, it would seem natural that the latter rise more steeply, otherwise excessive gamblers would be happier than non-gamblers. Accordingly, there must be a critical level of the consumption capital, \( C^* \), occurring at \( t^* \), beyond which the marginal costs of gambling exceed the marginal benefits (Note 1). Furthermore, there is a point in time, \( t^{**} \) (with \( t^{**} > t^* \)), at which the cumulative net benefit of gambling incurred during the period \( t=0 \) to \( t^* \) is outweighed by the net costs incurred beyond \( t^* \). In other words, \( t^{**} \) is the point at which the net benefit of gambling career turns from positive to negative. Figure 1 below sketches the course of the cost and benefit per period curves over time for a stereotypical pathological gambler (Note 2).

![Figure 1. A Pathological Gambler’s Cost and Utility of Gambling Per Period and Over Time](image-url)
There are four profiles of players for whom gambling can be rational according to this model:

1) Gamblers who are resistant to addiction

If the equilibrium level of consumption capital is attained before reaching the critical level ($C^* < C^+$), then this player gambles rationally; he may be referred to as being resistant to addiction (Note 3).

2) Only marginally addicted gamblers

For marginally addicted gamblers, their $C^*$ only slightly exceeds their $C^+$. In this case, the net benefit experienced in the first periods erodes too slowly beyond $t'$ to fully vanish before $T$.

3) Gamblers with a short time horizon

If a player has a short time horizon, specifically $T < t^+$, he does not experience the large net costs in later periods. For example, the player’s expected remaining life might be quite limited (Note 4).

So far, all three types of players manage to retain a net benefit from gambling, just as stipulated by inequality (7). Yet, there is a fourth profile to which this does not apply. It represents a departure from types 1) to 3): Having so far implicitly assumed that gamblers are perfectly informed about all aspects relevant to the gambling decision, I now relax that assumption.

4) Gamblers who made an initial error of judgment (but could not have known better)

Under imperfect information, (prospective) gamblers must make their decisions based on estimates relating to all the variables and parameters of the above model, i.e., based on expected values. These estimates will often prove to be wrong. Let us initially assume that the estimation errors are random, i.e., the estimates are correct on average. In particular, gamblers must form an estimate of their individual risk of addiction, which derives from several components of the model. Some potential gamblers will overestimate their risk and thus play too rarely or not at all, losing some utility they could have enjoyed. Others will be correct in their estimate and behave accordingly. The third and most interesting case refers to players who underestimate their risk of addiction. They will begin and continue to gamble even though their cumulative net benefit from doing so eventually turns negative. Their current gambling is not irrational because it is the logical consequence (path dependency) of an initial decision based on imperfect information, a decision that ex post proves to be unfortunate, rather than wrong.

Crucially, from such random mistakes the non-random or systematic misestimation of relevant arguments in the decision to gamble must be distinguished. In the subsequent sections, I will attribute such systematic mistakes to cognitive distortions. They imply that the players could, to a certain extent, anticipate their mistakes and act accordingly. It is their failure to not correctly respond to their cognitive distortions which makes their gambling irrational. Such irrational gambling is detrimental to the players’ welfare due to their (long-term) preference condition (expression 7) failing to coincide with their (short-term) participation condition, which is the same expression but with the player’s estimated values instead of the “true” ones.

Scope for a systematic bias in the participation condition exists in all aspects of the model. I will cluster these factors into three sources of irrationality to provide some structure for the subsequent survey of the available theoretical and empirical evidence:

1) Time preferences—refers to a systematic but temporal change to the discount function $D_t(t)$ while gambling, leading to an accelerated discounting of future events

2) Time horizon and frequency of play—refers to a systematic underestimation of the time horizon ($T$) of the decision to gamble and/or a systematic underestimation of the frequency of play

3) Costs and benefits—refers to a systematic underestimation of the costs of gambling and/or an overestimation of the benefits, thus involving the pleasure from gambling ($\alpha$), the pleasure from
winning (γ), the odds of winning \( P_{W} \), the size of the potential win (contributes to Ψ along with \( P_{W} \)), the costs of addiction (k), and the development path and equilibrium level of the consumption capital \( (C(t)/C^*) \).

3. Theoretical and Empirical Evidence of Irrational Gambling Decisions

Evidence of irrational gambling typically takes the form of measured differences in decision behavior between gamblers and non-gamblers or between gamblers in their “hot” state and gamblers in their “cold” state. While some of these differences clearly fall within the realm of what was defined above as irrationality, in other cases, that judgment is far less clear.

Such differences can be ascertained in two ways (Clark, 2010). First, the psychobiological approach examines changes to the subjects’ bodily functions, in particular their brain activity, in response to gambling stimuli. I will briefly summarize the somewhat limited evidence pertaining to that approach below. Second, the cognitive approach, which is much more prevalent, focuses on gambler’s outward behavior. A comprehensive review of existing evidence on cognitive distortions follows in the subsequent subsections.

Within the psychobiological approach, a body of research has detected deviating brain activity among pathological gamblers in the regions of the brain that regulate reward and emotion, such as the ventromedial prefrontal cortex (vmPFC), the striatum and the dopaminergic neurotransmitter system. Pathological gamblers have also been found to exhibit anomalies in various regions of the brain relevant to decision-making, in particular the (mis-)regulation of noradrenergic, serotonergic, dopaminergic and opioidergic neurotransmitters, as well as in genetic mechanisms such as the dopamine receptor D2 gene (Blaszczynski & Nower, 2007, p. 335). The vmPFC is responsible for a biased cost-benefit-assessment of addicts in general (Kalivas & Volkow, 2005) and of compulsive gamblers in particular (Bechara, 2003). During gambling (Reuter et al., 2005) and exposure to gambling cues (Potenza et al., 2003), vmPFC activity is reduced. A dysfunction in that region of the brain thus plays an important role in the development of gambling addiction, but also in other types of impulse control disorders (Brewer, Potenza, & Grant, 2007, p. 354).

This evidence is confirmed by results from studies relying on the Iowa Gambling Task, an instrument to measure the quality of decisions (Bechara et al., 1994). In the test, the subjects are presented two differently prepared decks of cards. One deck entails low rewards and low penalties, but yields long-term gains for the player. The other deck returns higher rewards and higher penalties; however, the player will lose in the long run. In this setting, decision quality correlates positively with the perfusion of the vmPFC and other cortical regions (Adinoff et al., 2003). Pathological gamblers (Cavedini et al., 2002) and generally individuals with a vmPFC disorder such as substance abusers (Bechara & Damasio, 2002) also show significantly lower results than the control group in this test, in that despite the greater risk, these persons tend to pick cards from the deck with the lower expected value (Note 5).

Similar results were obtained in a study by Regard et al. (2003) using a neurological approach: Pathological gamblers exhibit lower memory and mental performance, as well as lower performance in the executive functions (abstraction and planning). Furthermore, pathological gamblers show a distortion of attention with respect to gambling-related stimuli. Their mental response times when confronted with gambling-related words are significantly longer (Boyer & Dickerson, 2003), yet their style of reaction is impulsive in the sense of reduced reaction times in their decision behavior (Goudriaan et al., 2005). Both effects can contribute to decision-making deficits. Furthermore, such
gamblers exhibit deficits with respect to higher executive functions, delayed discounting and interference control, as well as elevated levels of impulsivity similar to that of persons suffering from attention deficit disorder (Bechara et al., 1994; Petry, 2001). Specifically delay discounting was shown to be related to problem gambling (Vanderveldt, Green, & Rachlin, 2015). Holden (2001) states that addiction in general sensitises the reward system in the brain to the “drug”. In consequence, the baseline release of dopamine is reduced (Bergh et al., 1997), strong rewards can increasingly only be obtained from the addictive behavior (Grüsser et al., 2007) and the addict “forgets” about other types of rewards (Grüsser & Albrecht, 2007, p. 22) (Note 6). Pathological gamblers thus incur a large cost in the form of their reduced pleasure from non-gambling activities—part of which they likely fail to internalize in their decision to gamble.

Pertaining to the cognitive approach, some general evidence of irrational gambling will very briefly be summarized before examining at greater depth the potential causes of such irrationality. Up to 80% of pathological gamblers’ verbalizations during the game are irrational, and such players will on average exhibit 3.5 different cognitive distortions (Toneatto et al., 1997; Gaboury & Ladouceur, 1989; Ladouceur et al., 1988). The latter team of authors conclude that a small number of wins suffice to produce irrational thoughts in the gamblers. Their study shows that the ratio of irrational to rational thoughts is five to one. The authors had 20 individuals play roulette, the game having been prepared in such a way that one group of players win 20% of the time while the other group win half the time. Despite this marked difference in outcomes, it was not possible to ascertain a difference in the players’ ratio of rational to irrational thoughts (Note 7). Meyer and Bachmann (2005, p. 100) find that a heightened state of excitement can lead to a narrowing field of attention and produce irrational thought, since less information is critically processed and fewer alternatives are considered, potentially leading to increased risk-taking.

Interesting as such general evidence of irrationality in gambling may be, more is to be gained—particularly with a view to policy implications—by reviewing more closely the body of evidence that points towards the specific sources or manifestations of irrationality. For that purpose, it is drawn upon the three-part structure derived from the model of gambling decisions.

3.1 Irrationality Type 1: Inconsistent Preferences
Steep discounting of future events is widely regarded as a source or irrationality, although it is unclear at which stage discounting becomes ‘steep’. Substance abusers generally discount more steeply than a control group (Ainslie & Haendel, 1983), and impulsivity causes pathological gamblers to exhibit similar decision deficits as alcoholics (Lawrence et al., 2009). Petry (2001a) showed that pathological gamblers have a significantly greater tendency to favor small immediate rewards over larger future rewards. This effect increases if gambling coincides with substance abuse. Related results are presented by Dixon, Marley, and Jacobs (2003) and Petry (2001b). Concurrently, substance abusers with gambling issues discount more steeply than those without such issues (Petry & Casarella, 1999). Impulsivity, compulsive gambling and substance abuse are thus correlated, though the direction of causality is not entirely clear. Results by Vitaro, Arsenault and Tremblay (1999) speak in favor of impulsivity as the cause, in that impulsivity in juveniles aged 12 to 14 is a good predictor of compulsive gambling at age 17. By contrast, the finding that former smokers and non-smokers discount less strongly than smokers (Bickel, Odum, & Madden, 1999) can be interpreted in both ways, given that smoking reduces life expectancy (Ross et al., 2008, p. 99). For pathological gambling, however, it may be assumed that causation emanates from impulsivity since gambling is not associated with shorter life expectancy.
A temporary shift of importance of near events relative to far events may be caused by an increased activation of the mesolimbic pathway (McClure et al., 2007, p. 5796), which may in turn be checked by the prefrontal cortex as a “higher supervisory institution” (Berridge & Robinson, 2003). The more strongly the prefrontal cortex is developed, the more likely it will prevail over the desires originating from the mesolimbic pathway. The result, consumption or no consumption, thus depends on the relative strength of the two systems (Bernheim & Rangel, 2004, p. 1563) and their relative current levels of activity. For example, a temporary, artificially induced boost to the mesolimbic pathway will promote short-term thinking and thus the tendency to seize immediate rewards to the detriment of larger future rewards. The consumer will later regret such a decision as soon as the balance between the two systems is once again restored.

While the prefrontal cortex tends to act slowly, the mesolimbic pathway is able to “fire” at very short notice. So the more rapidly a decision must be taken, the greater the importance of the mesolimbic pathway becomes in that decision (Bernheim & Rangel, 2004). Except for the case of lotteries, gambling decisions must typically be taken quickly, and discounting of future events temporarily increases in consequence. The importance of the mesolimbic pathway furthermore increases with the individual’s state of excitement. Gambling characteristically generates excitement (Petry, 2005, p. 121f) and thus, via the mesolimbic pathway, promotes short-term thinking. This applies not just while actually gambling but already when the player is exposed to various gambling cues to which he was conditioned by prior gambling. The excitement of a pathological machine gambler thus already rises when he sees or hears the machine; theses cues take him from the so-called ‘cold mode’ to the “hot mode”. Bernheim and Rangel (2004, p. 1559) define the latter as a state of heightened excitement in which consumption will always occur. According to Loewenstein (1996), the utility function is misinterpreted in the “hot mode”, while in the “cold mode”, all alternatives are taken into account and the utility function is maximized. So the more pronounced the artificial, gambling-induced state of excitement and the more quickly gambling decisions must be taken, the more dominant the player’s short-term thinking becomes. The potential for irrational decisions increases. Sobottka (2007) even attributes pathological gambling exclusively to the increased importance of the players’ mesolimbic pathway.

A major manifestation of instable time preferences and, thereby, a cause of potentially irrational gambling is impulsiveness. According to Ainslie (1975), impulsiveness is best modeled as hyperbolic discounting. In contrast to exponential discounting, hyperbolic discounting involves not a fixed but rather a variable discount rate, which is higher for shorter time horizons. By consequence, preference relations may shift as time passes, which potentially leads to a violation of the transitivity assumption of rationality. Imagine two rewards: Reward A materializes in period t₁ and the slightly larger reward occurs in period t₂. A discount rate that decreases over time can then imply that an individual prefers the larger and later reward B in periods t₀ to t₁ but switches to preferring the imminent but smaller reward A in t₁. Such preference reversal is not possible with exponential discounting.

Thaler (1981) provides empirical evidence of hyperbolic discounting. Subjects were asked how much money they require at a future date to be indifferent between that payment and a given sum today. If the future payment was scheduled in a month’s time, the subjects on average demanded an annualized interest rate of 345%. If, by contrast, that payment was to occur in ten years, the corresponding value was only 19%. Thus, the interest rate itself is discounted over time. Skog (2005, p. 124f) demonstrates the connection between hyperbolic discounting and addiction by transforming reward B into a disutility that results from the consumption of reward A. At t₀, the addict still feels that the benefit of consuming
3.2 Irrationality Type 2: Underestimating Time Horizon and Frequency of Play

Irrational gambling may ensue if the individual systematically underestimates the length of his consumption path, $T$. As a special case, he may incorrectly assume (or wish) that his consumption will end before period $t^*$, leading him to the wrong conclusion that his gambling career might have an overall net benefit.

Such non-consideration of the distant future, which stands in contrast to Becker and Murphy’s premises, is assumed in a model of tobacco consumption by Suranovich, Goldfarb and Leonard (1999). The authors depict juveniles as merely deciding how many cigarettes to smoke on a given day, rather than rationally planning their entire lives. The model thus assumes a very low degree of decision bundling. This form of bounded rationality finds strong empirical support (Fehr & Zych, 1998) and is easily transferred to gambling addiction. Success in resisting the temptation to gamble strongly depends on the number of consumption periods that the actor takes into account or how many individual decision steps he bundles in one fundamental decision. Skog (2005, p. 126) refers to the capacity to bundle decisions as “willpower”. Alternative economic models of addiction likewise emphasize the importance of strength of will (Schelling, 1978; Akerlof, 1991; O’Donoghue & Rabin, 1999).

Underestimation of the time horizon and frequency of play may arise from an erroneous estimate of the individual’s own future preferences, e.g., an underestimation of the future urge to gamble. According to Skog (2005, p. 122), such underestimation is regularly due to an information deficit. Preference forecast errors can for example be represented by the so-called ‘endowment effect’ (Tversky & Kahneman, 1991), confirmed empirically by Loewenstein and Adler (1995), which describes the fact that actors will attach a higher value to a good if it is in their possession compared to when it is not. This can be explained by the phenomenon that the welfare reduction caused by a given loss (convex function) exceeds the welfare increase caused by an equally sized gain (concave function). Surprisingly, individuals fail to anticipate this preference shift, even though it occurs within mere minutes. A possible explanation would be that the individuals forget their prior preferences (Marcus, 1986) and instead assume that their preferences have always been what they currently are, i.e., they are oblivious of the shift (Loewenstein & Adler, 1995). Skog (2005, p. 122) additionally shows that the human capacity to remember their needs is distorted due to an influx of dopamine that causes a positive memory bias. Consequently, addicts will often underestimate their urges and therefore underinvest in precautionary measures that can protect them from future consumption, i.e., so-called cue management. Similarly, a recreational player may systematically fail to take into account that his preference structure will change as he continues to play and his consumption capital accumulates.

Another fundamental cause of the underestimation of $T$ arises from the consideration of sunk costs in so-called mental accounts, an effect also known as “entrapment”. Entrapment means the holding on to prior decisions and continued investment, even though these decisions have already been proven wrong (Thaler, 1980). Individuals will form mental accounts for various affairs. If an account, e.g., for gambling, is overdrawn, the person will seek to rebalance it, even if that entails losses in other accounts and a decline in overall welfare. Gamblers thus tend to “chase” after their losses (Petry, 2005, p. 218f). Chasing as an expression of a person’s loss of control over their gambling behavior is so characteristic of gamblers that it constitutes a diagnostic criterion for pathological gambling and is regarded as an...
essential step in the development of gambling addiction (Lesieur, 1979). Moreover, chasing differentiates pathological gambling from substance addiction (Holden, 2001; Potenza, 2006). The longer, more frequent and more intense gambling that results from the urge to chase losses (Productivity Commission, 2010, 7.1) in turn leads to increased losses and, thus, to additional sunk cost effects. This loop intensifies as the duration of play increases (Beckert & Lutter, 2007). In a similar context, Breiter et al. (2001), using magnetic resonance imaging, examine how both humans and laboratory animals perceive gains and losses depending on a reference point. If the reference point changes, the reaction in the brain to an identical event also changes, meaning that the mental reaction depends on prior losses or gains. Thus, chasing is an expression of holding on to a reference point in form of the gambler’s prior wealth. It may also be an expression of social pressure. If a person changes his behavior in the face of indication that he should do so, he implicitly admits his prior mistake, which may constitute a large psychological cost to her. If he thus tries to fool herself, such behavior will satisfy most definitions of irrationality. And if his continued chasing is intended to fool others, he has likely succumbed to a misconception about the rationality of those around him (Skog, 2005, p. 123f.).

Regarding the prevalence of chasing, only between 0.2% and 1.7% of Australian recreational gamblers describe themselves as having one of a number of possible control issues (Productivity Commission, 2010). By contrast, the percentage for pathological gamblers lies between 64.3% and 87.5%. Wilcke (2013) empirically examines chasing among pathological poker players based on their actual playing behavior, distinguishing between within-session chasing (longer or more intensive gambling within a session following a loss) and between-session chasing (returning to the table after a short break, e.g. on the next day). The results demonstrate both significant between-session chasing and within-session chasing in the form of increased risk-taking for pathological gamblers.

3.3 Irrationality Type 3: Underestimating Cost, Overestimating Benefits
Unrealistic estimates of the expected value of future events are very common (Weinstein, 1980). Among the numerous factors that drive the expected costs and benefits of gambling, most of the available evidence in this context concerns the (estimated) probability of winning—arguably the variable in the model that is most prone to cognitive distortions. Before going into the details of the numerous reasons why and the ways in which gamblers overestimate the probability of success, a crucial distinction must be made: Ignorance of probability theory, as was already noted, is per se not associated with irrationality.

By contrast, scope for irrationality exists where gamblers know about the independence of random events at heart but at the same time somehow believe that the rules of statistics do not apply to their present situation. While gambling, their appreciation of these rules is temporarily suspended—a phenomenon known as “cognitive switching” (Sévigny & Ladouceur, 2004). Or in other words: “Knowing something and having this knowledge alter your behavior are often two different things” (Williams, Simpson, & West, 2007, p. 10).

The illusion of control is a particularly widespread cause of an overestimation of the expected pay-out among gamblers. The illusion refers to a gambler’s belief that he is able to achieve better results than probability theory would suggest (Petry, 2005, p. 210). Strickland, Lewicki, and Katz (1966) show, for example, that in dice games, subjects are willing to place higher bets if they (rather than other players) are allowed to roll the dice and to determine the winning numbers—even though the result is of course independent of the person (Note 8). The illusion of control is more pronounced the more choices a player has, the larger his degree of involvement in the game (e.g., pressing the buttons of a slot machine) and the shorter the time between placing the bet and learning the outcome (Grüsser &
Albrecht, 2007, p. 69f). The so-called “attributional bias” also forms part of the illusion of control. It refers to the tendency to overestimate the importance of situational factors in explaining events that are predominantly or wholly driven by chance (Petry, 2005, p. 210). For example, individuals will begin to believe in their forecast skills after correctly “predicting” a series of purely random events (Langer & Roth, 1975). Positive results are attributed to personal “skills” while negative results are blamed on circumstantial conditions (Gaboury & Ladouceur, 1989). This effect is stronger the greater the skill component of a game because a more obvious skill component is more easily overestimated.

The erroneous assumptions of being able to predict and to influence the results of the game are the most common cognitive distortions among machine gamblers (Delfabbro & Winefield, 2000). In an Australian study, 17% of recreational machine gamblers versus 28% of pathological gamblers strongly agreed that certain strategies exist which produce higher pay-outs (Productivity Commission, 2010). Given the right strategy, gambling can yield a profit in the long term—this expression of the illusion of control was shared by 8% of recreational gamblers across all types of games, by 25% of players at risk of addiction, and by even 32% of problem gamblers.

These results prove, on the one hand, that the prevalence of cognitive distortions increases with the risk of gambling addiction, as does, by implication, the degree of irrationality and, consequently, the proportion of the private costs of gambling that is relevant to welfare. On the other hand, given that the prevalence of cognitive distortion is not limited to pathological gamblers, part of the private costs of non-pathological gamblers, too, becomes relevant to welfare.

Next, evidence on the three major causes of a lacking appreciation of statistics is presented: illusion of causality, representativeness heuristic, and failure to understand the law of large numbers. I acknowledge the argument that a lack of understanding of statistics does not necessarily mean that someone is acting irrationally. Still, I believe it is quite likely that gamblers in their hot mode adhere less to statistical reasoning than in cold mode, which would mean they are act inconsistently and thus irrationally. In any case, not adhering to statistical reasoning can be interpreted as an information asymmetry, which also leads to the conclusion that part of the private costs are relevant to welfare and should be subject to regulation.

The illusion of causality refers to the assumption of a causal relationship between two unrelated events (Petry, 2005, p. 217). For example, machine gamblers may think that the winning probability depends on the size of the bet, the specific game played, the time of day, the day of the week or any number of other factors (Walker, Schellink, & Anjoul, 2007, p. 26). In consequence, the probability of winning is overestimated. The illusion of control may also be construed as superstition—a widespread phenomenon among gamblers.

The representativeness heuristic in turn describes the (mis-)estimation of the likelihood of a sample based on similarities in representation (Kahneman & Tversky, 1982). For example, most people consider a lottery ticket marked 3-4-5-6-7-8-9 less likely to win than one marked 8-13-18-24-27-33-39 (Petry, 2005, p. 214). The so-called “gambler’s fallacy” is also part of the representativeness heuristic. It refers to the mistaken belief that deviations from the mean in one draw will tend to be “balanced” by an opposite deviation in the next draw. For example, Corney and Cummings (1985) were able to show that the majority of people will predict “tails” for an impeding throw of a coin if previous throws returned heads-heads-tails-heads-heads-heads-heads, even though the events are, of course, independent.
If gamblers feel that such a “balancing” is due to happen, they tend to raise their bets (Meyer & Bachmann, 2005, p. 98). A gambler can also become prone to the opposite of the gambler’s fallacy: the “hot hand cognitive illusion” (Gilovich, Vallone, & Tversky, 1985), which likewise leads to increased bets. The hot hand fallacy refers to the overestimation of the probability of a streak continuing (rather than “balancing”). For example, 50-70% of Australian machine gamblers stated they believe that wins and losses appear in streaks, so machines can be “hot” or “cold” (Productivity Commission, 2010). Between 45% and 60% of players consider it disadvantageous to approach a machine that has just disbursed a win. Again, addicted gamblers are more prone to this distortion: Only 5% of recreational gamblers (across all types of games) believe that the winning probability increases with the length of a prior losing streak, while 20% of problem gamblers and 33% of pathological gamblers succumb to this erroneous notion.

Base rates also form part of the representativeness heuristic. They denote the frequency with which an event occurs within a population, which gamblers often fail to take into account in their decision making (Petry, 2005, p. 216). According to Tolkemitt (2000), the overestimation of very low probabilities, as displayed by most people (Kahneman & Tversky, 1979), can also be subsumed under the representativeness heuristic. The most important reason for this overestimation is that the probability of an event occurring makes little difference to the mental ability to picture that event. This phenomenon is particularly apparent when jackpot winners are presented in the media. Barseghyan et al. have empirically demonstrated this overestimation of low probabilities in the context of insurance (Barseghyan et al., 2013). The representativeness heuristic accordingly also leads to an overestimation of the expected value of gambling.

Closely related to the representativeness heuristic is the lacking appreciation of the law of large numbers, which says that sufficiently large samples are representative of the population (Tversky & Kahneman, 1971). Individuals often draw inferences about the population too early and from too small a sample (Petry, 2005, p. 215f), leading to a misjudgment of the chances of success in gambling. According to Turner (2000), pathological gamblers exhibit less understanding of random events and are therefore more likely to misjudge the underlying probabilities. The problem of misjudgment based on small samples is aggravated if the dispersion of the results is high because to obtain a given level of confidence, the sample must be larger in more widely dispersed populations. A high degree of dispersion also promotes the belief in patterns that are interpreted into a “random walk”. Such beliefs are particularly prevalent in games whose expected value is close to the bet size, as in roulette or other games with high pay-out ratios.

Also closely related to the representativeness heuristic, the availability heuristic refers to the imaginability of an event and the effect it has on its estimated probability (Petry, 2005, p. 216). For example, the risk of a tornado is typically overestimated, whereas the risk of diabetes is underestimated (Dowling & Chin-Fang, 2007). Tversky and Kahneman (1973) confirm the availability heuristic with the following experiment: Subjects were asked whether they thought that specific consonants appear more often as the first or as the third letter within a word. Most subjects (105 vs. 47) opted for the first position with the opposite being true. The reason is that a letter is more easily imagined in the first than in the third position. The effect is exploited by gambling operators who in their marketing show winners but never losers.

Petry (2015, p. 215) lists three sources of the availability heuristic. First, familiarity. Gamblers are more likely to play a game they know, even if this game carries a lower expected value than others. For example, a German sports better will likely choose soccer for his bets, and within that sport, he will bet
on the home team rather than the foreign team. Second, recency is important because memories fade over time. Petry (2005, p. 215) provides the example of disaster insurance: Sales of insurance policies are highest after a catastrophe has occurred and decline afterwards. Third, vividness means that unusual, rare or especially dramatic events are best remembered and are therefore considered more likely than mundane events. For instance, a victim of burglary will tend to overestimate the probability of another break-in. In the gambling context, this means that large wins are better remembered than many small losses, so the likelihood of the former is overestimated. This is particularly relevant for games with a jackpot, in which marketing measures can easily single out the winner from the mass of losers. It also follows that reinforcement need not occur via a player’s own reward but it is also possible indirectly by observing the winnings of others as second-order reinforcers (Petry, 2005, p. 202).

The availability heuristic is reinforced by so-called “decision framing”, which means that different portrayals of the same facts can produce different preferences (Tversky & Kahneman, 1986). For instance, the fire brigade will look better in a news coverage reporting that 80 of the 100 inhabitants of a burning house have been saved, compared to a report that they had to leave 20 of the inhabitants to die in the fire. Decision framing is an integral component of gambling. In a sense, it occurs automatically because providers quote the chances of winning but not the risk of losing, potentially leading to welfare-reducing decisions in those who succumb to this bias.

A particular form of the availability heuristic is triggered by so-called “near wins”, e.g., if a scratch card returns two of the required three identical symbols. Near wins are also second-order reinforcers (Petry, 2005, p. 202) and increase the motivation to play (Clark et al., 2012). Most notably, slot machines are programmed to produce near wins more often than pure chance would predict, as was first shown in 1966 (Strickland et al., 1966; Schüll, 2012). Gamblers tend to mentally transform such near wins into wins (Skog, 2005, p. 122), leading to strongly distorted conceptions of the pay-out rate. Harrigan (2007), for example, observed that machine gamblers overestimate the pay-out rate by a factor of two to five if they are able to see the movement of the wheels and the symbols above and below the payline. The phenomenon of near wins is also very prominent in sports betting where an anticipated win turns into a loss only at the last moment. Near wins have been extensively researched. They motivate gamblers to continue playing (Clark et al., 2009) and thus promote excessive gambling (Clark, 2010). While in the brains of non-gamblers they produce the same reaction as a loss, in gamblers’ brains the reaction is the same as to a win (Habib & Dixon, 2010; Chase & Clark, 2010). Slot machine gamblers show a particularly strong reaction to near wins (Productivity Commission, 2010).

Overall, about 60% of machine gambling revenues derive from players who hold beliefs that are incompatible with the laws of statistics (Productivity Commission, 2010, 4.36f). Interestingly, the share of revenues at gambling machines coming from addicted gamblers is similarly high, at 61% to 62% (Williams & Wood, 2007).

4. Discussion

4.1 Picoeconomics: A Reconciliation of Rational Choice and Behavioral Economics

The preceding survey of the vast body of existing theoretical and experimental evidence made it clear that gambling behavior is in several ways incompatible with the standard model of rational decisions. With regard to compulsive gambling, a consistent theoretical framework is missing to explain how it can always be rational. At best, it should be referred to gamblers as being “boundedly rational”, where the degree of boundedness—or the degree of irrationality—increases along a continuum with the
consumption capital, or with the level of addiction. Since consumption of gambling shares similar characteristics with the consumption of other addictive goods and behaviors, it can be argued that the model of rational consumption is not applicable to compulsive consumption.

Does this mean that, with respect to addiction, we must altogether reject the assumptions of rationality with all its analytical benefits? While classical microeconomics regards humans as consistent and rational actors, the subdiscipline picoeconomics, which Ainslie (1992) established, represents humans as sequences of actors (Ross et al., 2008, p. 62) and thus models several interacting actors within an individual. This idea allows to preserve the assumption of rationality if only it is applied to the sub-individual level. It is only at the level of sub-individual decision-makers that commitment strategies, as proposed by Schelling (1984), can be sensible.

In the simplest case, an individual’s decisions are assumed to be the outcome of the interaction between two sub-individual actors, who behave rationally and in accordance with microeconomic theory. The actors may have disparate preference functions, potentially making them antagonists, and they maximize their welfare over their individual lives, which need not coincide with the life of the person. So while both actors maximize consistent preference functions, the person’s resulting actions may well be inconsistent and thus appear irrational. The strategic interaction between the sub-individual players may assume a non-cooperative structure (such as a prisoner’s dilemma) that precludes an optimal (i.e., welfare maximizing) decision on the individual level; in other words, the struggle between the multiple agents translates into an intra-personal externality and thus a welfare loss to the individual.

The most relevant application of the picoeconomic perspective in the context of addiction is the decision between smaller sooner rewards (the pleasure of consumption) and larger later rewards (not having to bear the consequences). Different sub-individual actors have different preferences regarding these types of rewards and, depending on who prevails, the consumer will or will not exhibit impulsive behavior. For example, the frequency with which smaller sooner rewards are favored constitutes an important differentiating factor between recreational and pathological gamblers (Cavedini et al., 2002). This perspective thus permits us to explain the failure to bundle hyperbolically discounted rewards, which appears as inconsistent and thus irrational behavior on the individual level, without violating the rationality assumption on the level of the (sub-individual) deciders (Ross et al., 2008, p. 102).

4.2 Policy Implications

The inapplicability of the model of rationality to compulsive consumption poses two problems: First, it cannot be relied on the behavioral predictions that rationality would entail: predictions based on the assumption of rational consumption are inaccurate for addictive goods. Second, in light of the above evidence, the implications of rationality for consumers’ welfare and, in turn, for social welfare, cannot longer be taken for granted. Regulation might thus not only take into account the external costs associated with excessive consumption patterns but also part of the private costs.

If rationality prevails, consumer and producer rents can never be negative. If, however, humans do not act fully rationally on the level of the individual, intra-personal externalities exist. In the former case, cost-benefit analyses can ignore private costs for being fully internalized; regulation should then only be concerned with external costs. In the latter case, the intra-personal externalities provide an additional rationale for regulation.

Depending on the degree of rationality, intra-personal externalities may constitute a considerable share of total externalities, challenging the current regulatory regimes, which are mainly concerned with inter-personal externalities. This result suggests a shift in the regulation of strongly addictive consumer goods like tobacco, alcohol, gambling, sugar, and trans-fats. Regulation in these areas should then be
concerned with reducing irrational demand, for example by reducing or banning advertising. Another important implication is even stronger protection of juveniles to avoid irrational demand leading to path-dependent addiction. Furthermore, taxation of such consumption should be set to a level that reflects not only external costs but also the part of private costs that fails to be internalized, so as to reduce demand to the level that would be optimal given rationality. The prevalence of addicted consumers could provide an indication as to the appropriate tax level, with more strongly addictive goods receiving a higher tax increase than less addictive goods. Since addicted consumers are less price sensitive than non-addicted consumers (Productivity Commission, 1999; Clarke, 2008) this comes with the additional benefit of fulfilling the Ramsey criterion, which states that the level of taxation should be inversely related to a product’s elasticity of demand (Philander, 2014; Ramsey, 1927).

5. Conclusion
This article has suggested that rationality is a continuum of the internalization of private costs and benefits. Akerlof and Shiller (2015) introduced the term “phool” to refer to people who fail to act in their best interest and are instead misled by manipulation and deception. Thus, a phool does not act fully rationally, and it is not by coincidence that this type of actors was conceived of with slot machine gamblers in mind.

Phools can be of two types: informational and psychological. Informational phools act on insufficient or incorrect information. Their existence is widely accepted and regulators try to protect them by reducing either informational asymmetry or its consequences. By contrast, regulators still often ignore psychological phools, who either act on emotions overriding cognitive control in a temporary “hot mode” or on cognitive biases that—similarly to an optical illusion—lead them to misinterpret reality (Akerlof & Shiller, 2015).

This article lends strong support to the existence of both types of phools. Within the framework of a mathematical model of addiction, the theoretical and empirical evidence presented for the example of pathological gamblers suggests that addicted consumers do not fully internalize their private costs and benefits. The sources of their irrationality can be grouped in three broader categories: (1) inconsistent preferences, (2) underestimation of the time horizon and the frequency of consumption, (3) underestimation of costs and overestimation of benefits of consumption. Hence, many compulsive consumers are not fully rational and match the characteristics of informational and/or psychological phools.

This evidence is not necessarily inconsistent with rational choice theory. The picoeconomics perspective suggests the existence of multiple, fully rational sub-actors within each individual whose competing interests can lead to inconsistent and thus seemingly irrational choices on the aggregate individual level.

Still, irrationality in the sense of non-internalized private costs (or benefits) can be interpreted as a source of inefficiency and can thus constitute a rationale for regulatory intervention. Any cost-benefit analysis of gambling should therefore also consider those parts of private costs that gamblers fail to internalize, or else false policy implications may be deducted—a conclusion which may well apply also to other addictions. The crucial question for scientific analysis and policy-making alike is whether we should consider rationality the rule that is to be qualified only for certain types of behaviors and individuals, or whether the analysis of consumer welfare must eventually accept irrationality as the norm—and, if so, what the analytical consequences will have to be.
References


Published by SCHOLINK INC.


Notes

Note 1. $C^*$ may be equal to zero.

Note 2. It might seem sensible to play until $t^*$ and then stop. But at that point, the costs of prior gambling still persist while there are no more benefits, yielding a situation that is inferior to not starting to play in the first place.

Note 3. For this type of player, the cost and benefit curves will deviate from those depicted in figure 1: His costs per period never exceed the benefits.

Note 4. This case must crucially be distinguished from myopia. Players with a short remaining life span (small $T$) and myopic players similarly attach a value of (close to) zero to future costs and benefits, but for different reasons. Myopic players are characterised not by a small $T$ but by excessive discounting, $D(t)$.

Note 5. An interesting variation of this study would have the higher-risk desk return a higher expected value. The subjects with a vmPFC disorder would of course continue to pick the riskier deck, but at which point would the control group also switch to that deck?

Note 6. Note that this stands in contradiction to Becker and Murphy’s (1998) assumption that the utility derived from alternative goods is constant.

Note 7. This study may be criticised, however, for basing the distinction between rational and irrational thoughts on the individuals’ utterances regarding their thoughts – a controversial methodology (cf. e.g., Nisbett and Wilson, 1977).

Note 8. This higher price could be interpreted to reflect the additional pleasure from gambling that the player experiences when rolling the dice himself.