

THE BEHAVIORAL ECONOMICS AND NEUROECONOMICS OF DISORDERED GAMBLING:

A POLICY-FOCUSED SURVEY OF RESEARCH



National Centre for
the Study of Gambling



National
Gambling Board
a member of the dti group

SOUTH AFRICAN
SARGT
RESPONSIBLE GAMBLING TRUST



This report was prepared for the South African Responsible Gambling Foundation by Don Ross of the University of Alabama at Birmingham and University of Cape Town, David Spurrett of the University of KwaZulu-Natal and Rudy Vuchinich of the University of Alabama at Birmingham

TABLE OF CONTENTS

1. Gambling, Addiction and Behavioral Economics.....	10
1.1. Gambling in scientific focus.....	12
1.2. Behavioral economics	26
1.3. Behavioral economics of substance use and addiction	32
1.4. The matching law and temporal discounting.....	36
1.5. Temporal discounting, substance use, and addiction.....	42
1.6. Three BE theories of addiction	45
1.7. Extensions of BE research to gambling.....	54
1.8. How many kinds of disordered gambling are there?	58
2. Behavioral Economics, Impulsivity, Addiction and Disordered Gambling	64
2.1. Impulsivity and addiction	64
2.2. Discounting: delay and probability	71
2.3. Impulsivity and Disordered Gamblers.....	74
2.4. Conclusion.....	86
3. The Neuroeconomics and General Neuroscience of Pathological Gambling	87
3.1. Neuroeconomics	88
3.2. The Reward System.....	97
3.3. Neuroscience of addiction	118
3.4. Gambling as elementary reward system manipulation	130
3.5. The neuroscience of pathological gambling.....	135
3.6. Neuropharmacology of pathological gambling	144
4. Conclusions	150
4.1. General implications	152
4.2. Personal- and community-level interventions recommended by BE	156
4.3. Future behavioral economic research and NADGs	160
4.4. Interventions for AG	163
5. Appendix: Disordered gambling and gender.....	171

EXECUTIVE SUMMARY

The following pages consolidate the executive summaries offered at the beginning of each chapter.

Chapter One

The principal reason that gambling policy is both controversial and difficult to normatively frame is that gambling is irrational, in a specific sense. Gamblers both want to win each bet and to participate in an activity that is stimulating because there's a serious probability of losing. Thus when the state licenses commercial gambling operations and extracts significant revenue from it, many think this amounts to exploiting the weakness of some citizens. This is of moral import particularly because non-trivial numbers of people gamble to the point of causing severe distress, and sometimes even catastrophe, to themselves and their families. Some extreme problem gamblers – so-called pathological gamblers – are popularly taken to suffer from a clinical disorder in the same sense that drug addicts and alcoholics are taken to suffer from such disorders. It troubles many people that commercial gaming institutions, and the state, may be earning profits and revenues by to some extent preying on sick people.

There is only one general way to try to make progress with respect to confused and conflicted intuitions on the proper goals and norms of policy around problem gambling. This is to conduct scientific research that attempts to understand what motivates people to gamble, and why some people find it difficult or impossible to gamble within controlled and reasonable limits.

One immediate problem in scientific study of gambling is uncertainty over basic concepts. This uncertainty makes it difficult to coordinate on instruments for recruiting and screening research subjects and interpreting research results. The psychiatric profession has operationalized the concept of 'pathological gambling', but not milder forms of problem gambling that are sometimes thought to be stepping stones to pathological gambling. 'Disordered gambling' (DG) is used to refer to problem and pathological gambling together. DG cannot be operationalized unless problem gambling is operationalized.

Yet another important issue, related to these, is whether some, many or most pathological gamblers (PGs) are genuinely addicted, in a medical sense, or whether common talk of 'gambling addiction' should be understood as metaphorical. This matters because it is important to know whether what has been learned about the effective treatment of substance addicts can be applied directly to treatment of some, many or most PGs.

A recently emerged scientific discipline, Behavioral Economics (BE), studies the relationships between motivation and reward. BE researchers are especially interested in addiction and addiction-like compulsive behavior because it is the clearest instance, and perhaps the most important instance for policy, of behavior in which people repeatedly act against what they say are their preferences, and then regret their actions. This is not cheap talk: people spend significant resources trying, often unsuccessfully, to restrain their own behavior with respect to drugs, gambling, and some other objects of frequent preference reversal.

BE researchers have developed rigorous and well-confirmed theories of this sort of behaviour. These theories have been tested on both humans and other animals and found to predict and explain the data well. A principal finding of BE is that the normal behavior of people and other animals when they are choosing between immediate and delayed rewards in which the later reward is larger than the sooner one is to reverse preferences twice over time: once as the smaller reward comes close, and then again after they have taken that reward. Thus the behaviour of addicts and DGs is, in this central respect, normal. On the basis of this discovery, our main question turns around. Instead of asking 'Why do some people become addicts or DGs?', we must ask 'How do most people avoid becoming addicts and DGs?'

BE has supplied a general answer. People who are not addicts or DGs bundle their rewards by means of personal rules. A personal rule is imposed on a person by herself to restrict her access to smaller, more immediate rewards. The reason people don't just ignore their personal rules under temptation is that breaking a personal rule ruins it, and does so at the time the choice is made. This is an immediate cost, not a delayed one. It implies loss of all the later, larger rewards that depended on the personal rule. Thus these rewards are 'bundled' together in the present, instead of being scattered separately through the future.

Based on the BE model of choice over time, we propose an operationalization of problem and disordered gambling. We cannot state this without anticipating something for which we present evidence in Chapter 3, that some, many or most PGs are genuinely addicted in a neuroscientifically specifiable and diagnosable way. Then our operational classification of DGs is as follows. A 'problem gambler' is anyone who (i) gambles often enough to be consciously aware of trying to develop personal rules to limit their frequency of gambling or their average amount gambled per session or both, but who (ii) regards these rules as inadequately effective because they regret their own gambling behavior on a regular basis, and who (iii) does not meet the psychiatrically accepted (DSM-IV) criteria for pathological gambling. Then a non-addicted disordered gambler (NADG) is any problem gambler or pathological gambler who is not an addictive gambler (AG). A disordered gambler (DG) is anyone who is either a NADG or an AG.

Evidence indicates that there is a significant difference between AGs and NADGs that is immediately relevant to treatment strategies and policies. AGs have distinctive personal (neurochemical) properties and are best helped by direct therapy aimed at these properties. NADGs, by contrast, are not picked out by distinctive personal properties. They are neuroscientifically normal people who are unusually vulnerable to bundling failures in environments typical of gambling venues. A principal focus of research that aims at reduction of the prevalence of NADG behaviour should be the circumstances under which casino and other standard gambling environments interfere with bundling. Once these circumstances are identified, changing them will have costs, to industry, to the state, and to non-DG consumers of gambling services. Determination of policy will require a combination of normative reflection and cost-benefit analysis. But this can only be usefully done once behavioral science learns more about the benefits that can be achieved by policies that facilitate improved bundling, and the costs of implementing these policies.

Chapter Two

Addicts to and abusers of a wide range of substances are known to be more impulsive than otherwise comparable individuals, in the sense that they have stronger preferences for smaller rewards delivered soon, compared to larger ones that arrive after some delay. Behavioral economists mark this distinction by saying that substance abusers and addicts discount 'more steeply' for delayed rewards.

Another form of discounting concerns not delay, but probability. A person is impulsive with respect to probabilities if their valuation of a reward declines relatively slowly as the likelihood of the reward arriving is reduced. This kind of discounting is less extensively studied than delay discounting, but is clearly of interest for the study of disordered gambling. The empirical studies of delay and probability discounting undertaken to date and including subjects with varying gambling severity (including problem gambling) do not establish clearly whether disordered gamblers are primarily impulsive about delays (like substance abusers) or whether they have distinctive responses to probabilities as well. Further experiments could settle this question relatively simply, and thereby provide a more precise behavioral characterization of disordered gamblers.

Behavioral measures of impulsivity are not the only approaches. That said, behavioral measures, based on experimental measurement, do correlate with gambling severity more successfully than more qualitative and psychiatric measures based on self-report and clinical appraisal. This does suggest that they relate more directly to the phenomena of disordered gambling than psychiatric categories.

Cognitive psychological assessments of decision-making, especially the Iowa Gambling Task, also separate disordered gamblers, who do poorly at them, from normal control subjects. It is not, though, completely clear what the Iowa Gambling Task measures, or how it relates to a number of other important tools for assessing decision-making. Further behavioral experiments could shed light on these questions relatively simply, and thereby enable two relatively isolated bodies of work concerning disordered gambling to be connected. Work in neuroeconomics (surveyed in the following chapter) also suggests ways of connecting the two bodies of work.

Finally, we know that various factors can temporarily influence decision-making, some by changing discount rates. Some individuals may be more vulnerable to these influences, and some environments (such as casinos) might be unusually endowed with them. Both possibilities are relevant to debate over policy regarding gambling and disordered gambling.

Chapter Three

Two branches of neuroscience, neurochemistry and the newly emerged discipline of neuroeconomics, have jointly achieved major breakthroughs over the past few years in the understanding of addiction. Neuroeconomics studies the behaviour of the brain as a calculator of the relative values of different possible and actual rewards that a person or other animal could pursue. Neuroeconomists have developed a rigorous and well-confirmed model of the computation of reward values, known as the 'reward learning / predictor valuation model'.



Principal evidence for this model has been gathered by scanning subjects' brains under functional magnetic resonance imaging (fMRI) while they perform behavioural tasks under experimental control. This method has allowed the model to be tested in application to specific brain systems.

The most important of these is the so-called reward system, implemented by the joint action of an old part of the brain, the limbic system, and a part that has expanded with human evolution, the prefrontal cortex. This system integrates four main computational activities: (i) learning environmental cues that predict reward, (ii) learning comparative values of rewards, (iii) focusing attention on cues that predict rewards, and (iv) motivating the system to act on the basis of these cues. A finding of neuroscience that is surprising to common sense is that pleasure is only indirectly connected to reward, and is a relatively weak motivator of behaviour compared to other things, especially prospects for positive surprise.

Neurochemical and neurodynamical study of the reward system has broadly revealed how it works. Key to its operation are relationships among three main brain chemicals, or *neurotransmitters*. These are dopamine, serotonin, and GABA. Each is favoured by dedicated receptors in different parts of the system, and the arrangement of these receptors creates natural pathways by which the four computational processes above are carried out and integrated.

Drug addiction has been identified as, at the neurochemical and neurodynamical level, a pathology of the reward system. Different drugs exploit different pathways, with that exploited by stimulant drugs (cocaine, amphetamines) being the simplest, and those exploited by alcohol and opiates being more complicated. But all share general properties. They create flooding of dopamine into a particular part of the reward system. This causes the system to focus attention on the environmental cues that predict this flood, which are those associated with self-administration of the drug. The system then learns to orient toward these cues and is motivated to act on them. At the same time, glutamate levels are increased in prefrontal cortex and modify its neural connections. The modifications in question interfere with normal prefrontal inhibition of older brain systems and thereby cause behavior to become more impulsive. This facilitates drug-taking, and so addiction amplifies and locks itself in. The reward system becomes less receptive to alternative sources of reward. As addiction solidifies, the system learns to respond not to the drug itself, effects of which are readily predictable, but to the cues that predict it.

The neuroeconomic model explains why this all happens. Because the reward system is a *learning* system, it responds mainly to *surprising* events. This aspect of the system is responsible for the familiar pattern whereby an addict becomes *sensitized* to her drug, consuming ever more of it and with increasing concentration on it to the exclusion of other rewards.

A good deal of recent evidence strongly suggests that gambling, in some people, triggers reward system response that mirrors its response to the most direct form of chemical hijacking, that of stimulant drugs. This evidence is reviewed in depth in the chapter.

In light of what we now know about the reward system, and in light of the neuroeconomic model, it is less surprising that gambling should be a target of addiction. Gambling activities are *engineered* to produce rewarding surprise – that is, to produce exactly what stimulates the reward system. The gambler is buying a maximally convenient, direct manipulation of her mesolimbic reward system. If the result is neuroadaptation of prefrontal circuits to reduce inhibition of impulse, her brain is changed into the characteristic brain of the addict.

She will experience cravings if she stops gambling, because other rewards will be less able to attract the attention of her reward system or motivate action. Thus she will be prone to relapse if she adopts a policy of trying to abandon gambling as its costs to her mount.

Fortunately, knowledge of the neurochemistry of addiction brings with it the promise of chemical interventions to counteract it. The chapter reviews very recent pilot studies – there being nothing more extensive as yet – that strongly suggest that AG can be effectively treated by drugs that disrupt addictive learning. These drugs do this by interfering with the neurotransmitter interactions that produce dopamine response amplification and sensitization. Of particular interest and promise are so-called ‘atypical antipsychotics’ that inhibit reuptake of dopamine in the heart of the reward system. The chapter concludes by indicating grounds for hope that within a few years there will be standard neuropharmacological therapies that allow addictive gambling to be effectively managed in the majority of cases, or perhaps even in all cases.

Chapter Four

There is only one general way to try to make progress with respect to confused and conflicted intuitions on the proper goals and norms of policy around problem gambling. This is to conduct scientific research that attempts to understand what motivates people to gamble, and why some people find it difficult or impossible to gamble within controlled and reasonable limits.

One problem in scientific study of gambling is uncertainty over basic concepts. This uncertainty makes it difficult to coordinate on instruments for recruiting and screening research subjects and interpreting research results. The psychiatric profession has operationalized the concept of ‘pathological gambling’, but not milder forms of problem gambling that are sometimes thought to be stepping stones to pathological gambling. Another important issue is whether some, many or most pathological gamblers (PGs) are genuinely addicted, in a medical sense, or whether common talk of ‘gambling addiction’ should be understood as metaphorical.

Based on the BE model of choice over time, we propose an operationalization of disordered gambling and all of its sub-varieties. A ‘problem gambler’ is anyone who (i) gambles often enough to be consciously aware of trying to develop personal rules to limit their frequency of gambling or their average amount gambled per session or both, but who (ii) regards these rules as inadequately effective because they regret their own gambling behavior on a regular basis, and who (iii) does not meet the psychiatrically accepted (DSM-IV) criteria for pathological gambling (PG). Then a non-addicted disordered gambler (NADG) is any problem gambler or pathological gambler who is not an addictive gambler (AG). A disordered gambler (DG) is anyone who is either a NADG or an AG.

There is a significant difference between AGs and NADGs that is relevant to treatment strategies and policies. AGs have distinctive personal (neurochemical) properties and are best helped by direct therapy aimed at these properties. NADGs, by contrast, are neuroscientifically normal people who are unusually vulnerable to bundling failures in environments typical of gambling venues. A principal focus of research that aims at reduction of the prevalence of NADG behaviour should be the circumstances under which casino and other standard gambling environments interfere with bundling.



A second class of recommendations from BE models are aimed more directly at NADGs and their significant others, the implementation of which would be a matter for individuals and thus would not entail significant industrial or regulatory reform. The gambling industry may, however, have an interest in participating in and facilitating the operation of an effective counseling system.

The alcohol treatment community and scientific treatment literature are much older and much more extensive than those for NADG. The alcohol treatment system has evolved from a focus on the unitary disease model and very intensive interventions to a focus on a multidimensional biopsychosocial model with a range of interventions that are appropriate for individuals with a range of severity of alcohol problems. Initial indications are that the nascent NADG treatment literature is attending to these developments.

Specific personal-level intervention recommendations based on BE are discussed. First, possible methods to improve NADG's ability to bundle future rewards are recommended. Second, we recommend several possible ways in which the price of access to gambling opportunities could be increased for NADGs, and several possible ways in which the price of opportunities for social interaction could be decreased for NADGs. Third, we recommend interventions aimed at addressing a particular perceptual distortion among NADGs regarding the value of gambles.

A wide range of behavioural research programmes could contribute significantly to our understanding of disordered gambling, especially among NADGs. Existing work establishes that BE has something to tell us about disordered gambling, but we don't yet know much of what that is. A range of desirable research programmes are briefly described, with remarks on what they could be expected to tell us.

Some, and more probably most or all, PG as operationalized by DSM-IV is in fact addictive gambling (AG). This implies a straightforward policy goal: measures should be taken to try to prevent AGs from accessing commercial or other organized gambling environments. We recommend that responsibility to deny AGs access to gaming establishments should be vested in the operators of the establishments. However, the reasonableness of this recommendation is dependent on resources being provided to make AG identification practical. At present, this condition is not satisfied. The two prospective scientific means of screening for AG, fMRI and brain fluid examinations, are absurdly expensive procedures for application to mass diagnostics outside self-reporting populations.

Two considerations rescue our policy reflections in this area from futility, however. First, evidence indicates that experienced casino staff detect PG customers with high reliability, and that the requisite level of experience is typically acquired quite quickly. Second, if most PGs were not AGs, then the ability of the studies we reviewed to statistically measure AG in populations drawn using screens that over-diagnose PG would be mysterious. Thus there are grounds to suppose that casinos are *already* able to identify and screen out AGs, except perhaps AGs in very early stages of morbidity.

The great hope encouraged by developments in neuroscience of AG is that we will soon be able to exclude AGs from gambling venues in the most benign and welfare-enhancing way of all: by means of a neuropharmacological technology that will reduce or eliminate AGs' interest in gambling in the first place, and their disposition to harm themselves if and when they do find themselves exposed to gambling environments for one reason or another.

However, until the neurochemically specified conditions for AG are determined to general satisfaction in the scientific and clinical communities, and then made the basis for sustained epidemiological research, efforts to estimate the costs to the gaming industry of fully, or nearly fully, managing AG are shots in the dark. Though we are sanguine about neuropharmacological research on AG continuing to be well supported, there is need for *integrated* research that includes BE. The recent breakthroughs in understanding addiction in general and AG in particular rested on the conjunction of basic neurochemistry and neuroeconomics. Neuroeconomics relies heavily on fMRI experiments which in turn require batteries of well-tested behavioral task protocols from BE.

At present there is little capacity for basic fMRI research in South Africa. However, there is capacity for supporting BE work of the kind we have identified as part of the next wave of research that, we hope, will soon turn AG from a serious social problem into a minor nuisance.

1. Gambling, Addiction and Behavioral Economics

EXECUTIVE SUMMARY

The principal reason that gambling policy is both controversial and difficult to normatively frame is that gambling is irrational, in a specific sense. Gamblers both want to win each bet and to participate in an activity that is stimulating because there's a serious probability of losing. Thus when the state licenses commercial gambling operations and extracts significant revenue from it, many think this amounts to exploiting the weakness of some citizens. This is of moral import particularly because non-trivial numbers of people gamble to the point of causing severe distress, and sometimes even catastrophe, to themselves and their families. Some extreme problem gamblers – so-called pathological gamblers – are popularly taken to suffer from a clinical disorder in the same sense that drug addicts and alcoholics are taken to suffer from such disorders. It troubles many people that commercial gaming institutions, and the state, may be earning profits and revenues by to some extent preying on sick people.

There is only one general way to try to make progress with respect to confused and conflicted intuitions on the proper goals and norms of policy around problem gambling. This is to conduct scientific research that attempts to understand what motivates people to gamble, and why some people find it difficult or impossible to gamble within controlled and reasonable limits.

One immediate problem in scientific study of gambling is uncertainty over basic concepts. This uncertainty makes it difficult to coordinate on instruments for recruiting and screening research subjects and interpreting research results. The psychiatric profession has operationalized the concept of 'pathological gambling', but not milder forms of problem gambling that are sometimes thought to be stepping stones to pathological gambling. 'Disordered gambling' (DG) is used to refer to problem and pathological gambling together. DG cannot be operationalized unless problem gambling is operationalized.

Yet another important issue, related to these, is whether some, many or most pathological gamblers (PGs) are genuinely addicted, in a medical sense, or whether common talk of 'gambling addiction' should be understood as metaphorical. This matters because it is important to know whether what has been learned about the effective treatment of substance addicts can be applied directly to treatment of some, many or most PGs.

A recently emerged scientific discipline, Behavioral Economics (BE), studies the relationships between motivation and reward. BE researchers are especially interested in addiction and addiction-like compulsive behavior because it is the clearest instance, and perhaps the most important instance for policy, of behavior in which people repeatedly act against what they say are their preferences, and then regret their actions. This is not cheap talk: people spend significant resources trying, often unsuccessfully, to restrain their own behavior with respect to drugs, gambling, and some other objects of frequent preference reversal.

BE researchers have developed rigorous and well-confirmed theories of this sort of behaviour. These theories have been tested on both humans and other animals and found to predict and explain the data well.

A principal finding of BE is that the normal behavior of people and other animals when they are choosing between immediate and delayed rewards in which the later reward is larger than the sooner one is to reverse preferences twice over time: once as the smaller reward comes close, and then again after they have taken that reward. Thus the behaviour of addicts and DGs is, in this central respect, normal. On the basis of this discovery, our main question turns around. Instead of asking 'Why do some people become addicts or DGs?', we must ask 'How do most people avoid becoming addicts and DGs?'

BE has supplied a general answer. People who are not addicts or DGs bundle their rewards by means of personal rules. A personal rule is imposed on a person by herself to restrict her access to smaller, more immediate rewards. The reason people don't just ignore their personal rules under temptation is that breaking a personal rule ruins it, and does so at the time the choice is made. This is an immediate cost, not a delayed one. It implies loss of all the later, larger rewards that depended on the personal rule. Thus these rewards are 'bundled' together in the present, instead of being scattered separately through the future.

Based on the BE model of choice over time, we propose an operationalization of problem and disordered gambling. We cannot state this without anticipating something for which we present evidence in Chapter 3, that some, many or most PGs are genuinely addicted in a neuroscientifically specifiable and diagnosable way. Then our operational classification of DGs is as follows. A 'problem gambler' is anyone who (i) gambles often enough to be consciously aware of trying to develop personal rules to limit their frequency of gambling or their average amount gambled per session or both, but who (ii) regards these rules as inadequately effective because they regret their own gambling behavior on a regular basis, and who (iii) does not meet the psychiatrically accepted (DSM-IV) criteria for pathological gambling. Then a non-addicted disordered gambler (NADG) is any problem gambler or pathological gambler who is not an addictive gambler (AG). A disordered gambler (DG) is anyone who is either a NADG or an AG.

Evidence indicates that there is a significant difference between AGs and NADGs that is immediately relevant to treatment strategies and policies. AGs have distinctive personal (neurochemical) properties and are best helped by direct therapy aimed at these properties. NADGs, by contrast, are not picked out by distinctive personal properties. They are neuroscientifically normal people who are unusually vulnerable to bundling failures in environments typical of gambling venues. A principal focus of research that aims at reduction of the prevalence of NADG behaviour should be the circumstances under which casino and other standard gambling environments interfere with bundling. Once these circumstances are identified, changing them will have costs, to industry, to the state, and to non-DG consumers of gambling services. Determination of policy will require a combination of normative reflection and cost-benefit analysis. But this can only be usefully done once behavioral science learns more about the benefits that can be achieved by policies that facilitate improved bundling, and the costs of implementing these policies.

1.1. Gambling in scientific focus

Except where cultures specifically set up institutional barriers to gambling, the majority of human beings are disposed to regularly stake wagers against each other on the outcomes of natural and contrived processes.¹ No one seriously doubts that one motivation of most people, most of the time they behave in this way, is enjoyment of the periods of uncertainty that fall between placing bets and seeing how they come out. When people win bets they typically get a further shot of satisfaction, since almost everyone enjoys becoming a bit richer, especially when this results from something they themselves chose to do and they can thus congratulate themselves.

To scientists, stating the obvious, as we just did, is typically the first step in exposing what we don't know and should take steps to find out. What proportion, if any, of gambles are mainly motivated by the lure of excitement, and what proportion are mainly efforts to get money? Are there other rewards that significant numbers of people pursue by gambling, such as feelings of control or escape from unpleasant realities? One cannot answer these questions just by reflecting on one's own case or that of one's acquaintances, since these samples aren't likely to be representative. Furthermore, psychologists have recognized for many years that people's reflective judgments about their own behavior are highly unreliable, because they more resemble story creation than objective reporting. The moment we recognize this, and make systematic efforts to collect and test samples of gambling behavior that *are* representative and objective, we're embarked on scientific study of gambling.

There are two main grounds for thinking such study worthwhile. One is the fundamental engine of science, basic curiosity. Many people want to know more about the general patterns of human behavior just because they find people (in general) interesting. It isn't *obvious* that nature would have designed people to like to gamble, yet the majority of them clearly do. Why? To the extent that we make progress on answering that, we automatically make progress on answering further interesting questions, because the explanation of why people enjoy gambling is bound to imply explanations of other things they like and dislike doing.

The other main motivation for scientifically studying gambling is the hope of furnishing a basis for accurate prediction of behavior. There are benefits in improving our ability to forecast when people will gamble, and how much they'll gamble. These benefits are obvious for someone who's in, or considering entering, the business of supplying people with gambling opportunities by bringing prospective gamblers together with things to gamble on. People who enjoy gambling themselves have an interest in predicting the behaviour of fellow gamblers. And it is a commonplace of both folk wisdom and scientific psychology that people usually need some knowledge of human behaviour in general in order to best predict what they themselves will do in different kinds of circumstances. A very inefficient way of finding out what sort of things you'll do under different conditions is to actually put yourself under all conditions and find out directly; a more sensible approach, at least at a broad level, is to suppose that you're typical of some reference class (i.e., women in general or South Africans in general or South African women or South African women in their forties, etc.) except where you have specific evidence to the contrary, and then find out by means of surveys what typical people in that reference class do.

¹ Gray, P. (2004). Evolutionary and cross-cultural perspectives on gambling. *Journal of Gambling Studies* 20: 347-371.

Let us give an example of this. The gambling industry, as noted above, makes money by supplying people with opportunities to gamble. Partly this just involves providing geographical focal points where people who want to bet on things can find other people to bet against. But it also involves supplying games on which to gamble. What should these games be? It is just 'obvious' to most people that they should be slot machines and card games and so on, because that is what they're used to. When they 'look inside' themselves they 'find' preferences for these sorts of activities.

But this seems like an explanation only because we haven't really confronted the full force of the question. Why do bookies and casinos get paid to supply people with things to gamble on in the first place? After all, every environment always has elements of uncertainty. People could bet on the weather, or on the colour of the next car that will pass, or on whether a squirrel will run up a tree or down it. Indeed, some people *do* enjoy betting on these sorts of things. But evidently most people like having a pre-established, structured common focus for their betting; otherwise the existence of *expensive* casinos, for which all gamblers pay a share of the costs, would be mysterious. And clearly people also like gambling on some *kinds* of uncertainties more than others, since casinos don't concentrate on finding games that cost as little as possible to set up, as they would if people were naturally indifferent and could be trained like laboratory rats to prefer whatever was provided.

Finally, almost all non-industrial cultures have converged on culturally preferred games with conventional rules for placing bets that require gathering and maintenance of some specific infrastructure (fighting cocks, cowrie shells, specially shaped pieces of bark, etc.).² Notice that once we seriously ask the question 'Why are people more attracted to certain sorts of gambling scenarios than others?' we're *approaching* the question of why they gamble at all; but as we saw, we can't make progress on this by each of us introspecting. We have to study lots of human behavior, and we have to do so by designing experiments that force nature to answer specific queries we put to it.

These considerations emphasize the value of better behavioral prediction for *personal* policies. Gambling also has a *public* policy aspect. Increasingly, many countries, provinces and municipalities get important revenue from supplying or licensing and taxing gambling opportunities such as lotteries and casinos. Thus we have an interest in being able to predict what different levels of public revenue we should expect from different possible institutional arrangements. Aiming at public revenue from gambling proceeds implies that we think it appropriate that there could be some transfer of a bit of the public finance burden, in which to some extent people who like to gamble pay higher per capita taxes than others, all else being equal.

But then who are these people that our gambling policy will lead us to tax more highly? Are they richer than average, randomly distributed in wealth or poorer than average? Are they older than average or younger than average? Are they more likely to be men or women? Better educated or less educated? In a diverse society like South Africa we should want to know if they'll be disproportionately drawn from one cultural community by comparison to another.

In asking these questions about who gambles more and less, we so far haven't distinguished gambling, in kind, from any other sort of behavior that's relevant to distribution and justice issues in public policy. We'd have reason to pose and to try to answer similar questions if we were setting out to design a new tax on bread.

² Gray, P. (2004). Evolutionary and cross-cultural perspectives on gambling. *Journal of Gambling Studies* 20: 347-371.

But gambling raises additional complications because it differs from bread-buying in two important respects: (i) no one regards the (normal) consumption of bread as *morally* worrying; and (ii) we have no reason to think that many people are importantly *irrational* in the way they consume bread.³

Many people think that gambling is morally improper. Others think that light gambling is morally harmless but that heavy gambling – that is, gambling beyond some level of frequency or with stakes that exceed some approximate threshold of someone’s resources – is morally improper. Still others think that gamblers themselves commit no moral offense, but that the state does something morally improper if it takes advantage of people’s enjoyment of gambling to get extra taxes from them. This belief might in turn have one or both of two different bases: one might think that the state shouldn’t exploit gamblers because gamblers are too likely to be economically vulnerable in other respects; or one might think that exploiting gamblers involves profiting from people’s irrationality, and that doing that is a form of preying on weakness, thereby resembling laying an extra tax burden on the physically or (in this case) mentally disabled.

Questions about whether certain kinds and patterns of gambling behavior, or certain public policies around gambling, are *in fact* morally bad, are philosophical questions. Questions about why people have the moral opinions they do, and about why and under what circumstances people do things they themselves think are immoral – thus opening themselves to immoral exploitation by their own lights – are anthropological and sociological questions. In this survey we will set both of these kinds of questions aside, though we don’t deny their importance. Readers who want to follow up on them in a careful way are referred to a recent discussion by Peter Collins.⁴ Here we will take one of Collins’s main conclusions on moral attitudes to gambling as our starting point. This is that, regardless of what one thinks is the *right* answer to the philosophical question, there is a clear majority *opinion* on this question that prevails at the moment in most democratic jurisdictions, including South Africa. This (several-part) opinion is that

1. gambling at moderate frequency for moderate stakes (as referenced to people’s own income levels) is something that people should be allowed to do if they wish;
2. the state should tax this activity at a rate comparable to the rate at which it taxes other forms of adult-focused entertainment that doesn’t enhance health; but
3. the state should regulate and keep within normative bounds the visibility and extent of gambling, should not allow exploitation by either itself or by private parties of people who can’t control their gambling, and should carefully monitor, with a view to managing, the extent to which gambling disproportionately draws resources from economically vulnerable people, especially children whose caregivers gamble.

³ This is a bit of a simplification. Some people, we might think, show irrationality by consuming so much bread and other starchy food that they become obese and harm their health and social lives. If we think we should try to help the irrational by manipulating their behavior, this could be the basis for an argument that taxes on food should be higher the less healthy it is. But we should recognize that most fat people who don’t attempt diets probably aren’t irrational in any important sense; they just like eating quite a lot, and value this highly relative to the goods they trade off for it.

⁴ Collins, P. (2003). *Gambling and the Public Interest*. Westport, CT: Praeger.

Again, let us be clear that we neither endorse nor criticize this consensus that Collins identifies. Some (especially religious) people are inclined to doubt its moral soundness and therefore work to change it. One of the foremost scientific authorities on gambling behavior, Howard Rachlin,⁵ has argued that, in light of the psychological factors that cause people to gamble, the current consensus doesn't lead to *stable* policy. That is, Rachlin thinks that commitment to norms (1) and (2) above almost inevitably leads governments to have problems maintaining coherent policies in line with norm (3), for two reasons. First, policies in keeping with norms (1) and (2) *cause* more people, according to Rachlin, to lose control of their gambling behavior. Let us call this 'the pessimistic behavioral hypothesis'. Second, Rachlin thinks that governments that depend on gambling for important revenues will be unable to avoid the temptation to exploit people who have trouble controlling their gambling because there is so *much* revenue in doing so, and these people don't have a strong lobby working on their behalf. Let us call this 'the pessimistic political hypothesis'.

The work we review in this survey will not bear on the pessimistic political hypothesis. That is a subject for political economy. However, our survey certainly will bear on the pessimistic behavioral hypothesis. We cannot claim to know whether liberal gambling policies contribute to the extent or severity of gambling problems except insofar as we have some idea why most people gamble, why some people gamble more than others, and why some people try to reduce or cease their gambling and find that they cannot (or, at least, that they cannot without special help).

It is the call for explanation of this last fact that confronts us with the supposed *irrationality* of some gambling behavior. We need to consider this idea carefully here, for two reasons. First, as we will see, the view that some or all gambling is irrational is crucial to the public policy issues around the legal status of gambling, the state's role in sponsoring it and profiting from its revenues, and the social obligations of the gambling industry. Second, it is because of the apparent irrationality of some gambling behaviour, in a particular sense we define shortly, that we can't study it using only the tools we bring to bear on other consumption behaviour, and must in addition resort to a distinctive scientific approach called 'behavioral economics' that we describe in section 1.2.

Let us begin with the role of irrationality in the public policy issues around gambling. Remember that we have put aside the philosophical question about whether gambling is *in fact* morally permissible, in favour of asking what kind of policy is best *given* the approximate liberal-democratic consensus on this question identified by Collins. The third part of that consensus says that the state should act to soften the harm done by gambling to society in general, trading this off against our aim to specially tax more rational gamblers. But no part of the consensus as we stated it here says anything about whether or to what extent the public authority should be obligated to help people who think that their gambling is a problem *to themselves*.

Modern democracies are, in general, deeply ambivalent and unsure about this sort of issue. On the one hand, most people shy away from too much *paternalism* – the view that some eccentric preferences people have are crazy in themselves and should be interfered with *over and above* the implications they have for the welfare of other people. For instance, most of us might think that handling cobras is mad. Yet a few people enjoy doing it. South African law tells them not to do it in circumstances where snakes might escape their control and menace others.

⁵ See Rachlin, H. (2000). *The Science of Self-Control*. Cambridge, MA: Harvard University Press.

However, a South African man recently made headlines by setting records for time spent living in rooms full of cobras and mambas, until his unsurprising demise at the fangs of one of his reptiles. The law did not try to interfere with him; and most South Africans probably think it was right not to have.

If we say that this person was irrational, what we mean is that *he wanted things out of life that we find strange*, but which were dangerous only to himself – and most of us seem to think that intervention against *that* kind of irrationality is not the business of the state.

But now suppose that the snake enthusiast had told friends and the news media that, fearing for his life, he was desperate to stop consorting with cobras, but found himself unable to follow through on his resolutions to do so. No matter how sincerely and fervently he swore off snakes, he reported, when the moment to leave them safely on the other side of some glass arose, he could not resist the impulse to get himself on the dangerous side of the partition. Many or most people of liberal-democratic views might in these circumstances think they had an obligation, or at least a *legitimate reason*, to help the unfortunate ‘cobraholic’ achieve *his own* stated goal in life by making it impossible for him to come within range of his siren’s call.

The different judgments in these two cases – the real one and the hypothetical one – stem from recognition of the fact that the cobraholic is irrational in a more profound way than was the late real man. The latter, as we saw, just had peculiar preferences. In the case of the imaginary more to do with snakes, but never took any actions consistent with this declaration, we might think he was just being disingenuous to keep his audience on edge. But suppose he took time and trouble and spent money to keep snakes out of his life, and then spent yet *more* time and trouble and money getting around these very barriers he’d set up for himself. *Then* we might judge either that ‘he doesn’t know what he wants’ or that ‘he really wants to stay away from snakes but can’t pull this off on his own’. In either case, we might think we could, and/or should, help him.

The case of the imaginary cobraholic is fanciful; the real snake man did his thing while observers shook their heads in bemusement and then met his predictable end. But it is easy for anyone to understand the made up case because it has close parallels in the situations of the many people who take actions to try to stop drinking or taking drugs or gambling but fail in their attempts.

Economists use the ideas of rationality and irrationality in a narrower sense than other people. To economists, only the more profound kind of irrationality exhibited by the cobraholic really *counts* as true irrationality. This isn’t because economists don’t recognize strange preferences, like the real snake man’s, as being strange. Instead, they begin by taking note of the same distinction everyone else does – between cases of strange but clear preferences and cases where people don’t seem to even *have* straightforward or consistent preferences – and then, for the sake of avoiding confusion in their professional discourse, agree to *call* only the second kind of phenomenon ‘irrationality’. This reflects more than just an arbitrary taste in the use of words. It stems from some deep facts about the relationship between the subject matters of economics and that of the neighboring science of behavior, psychology. The next section of the survey describes this relationship.

Before we go on to that, however, let us draw attention to another aspect of the imaginary case of the cobraholic that the reader will have noted. Irrationality as economists understand it corresponds to a core *part* of what we get at in everyday contexts by the concept of ‘addiction’. The historical origins of this concept are closely related to the concept of ‘enslavement’ which at one time was called by the same word. Gradually, ‘addiction’ came to mean ‘enslavement to an object or to a kind of behaviour’ (to what psychologists generalize as ‘a reinforcer’), as opposed to enslavement to a person.

Modern democracies have outlawed literal enslavement to persons.⁶ But there is plenty of ‘addiction’ around, and modern societies are less sure about how they should handle it. There may be no actual cobraholics (or maybe the real snake man was actually a cobraholic – we’ll come back to this question later, in the final chapter of this report). But there are certainly people who try to drink less or nothing but end of drinking a lot and severely harming themselves. And there are also significant numbers of people who try to gamble less, or with smaller stakes, or not at all, but end up gambling to ruin.

We will come back, briefly in section 1.8 and then at length in Chapter 3, to the question of how helpful (or not) it is to lump people who are irrational (in the economist’s sense) about gambling with people who are irrational about alcohol and drugs. For the moment, we put this aside in order to stipulate a more precise way to talk specifically about different sorts of gambling behavior. We’ll need a finer set of distinctions than just the bare one between people who have an unusually strong but not irrational (in the economist’s sense) taste for gambling and people who would like to change their gambling behavior but have trouble doing so. Psychologists have developed some conventions for separating kinds of cases here, and we need to follow these conventions, both to avoid confusions and so that we can measure the extent of different social phenomena with consistency from one survey or experiment to the next.

One of the most crucial steps in setting up a problem for scientific study is the definition of fundamental concepts. Carelessness here can, all by itself, predetermine or overwhelmingly influence the later interpretation of results and so make whole investigations useless. Let us give an illustrative example for the case of gambling problems. Suppose on the one hand that one adopted a very liberal definition, according to which anyone who had ever regretted any particular participation in a game or any particular bet was said to have a “problem” caused by gambling. In that case, everyone who has ever gambled more than two or three times would be declared to have a gambling problem, since almost everyone who has bet that often has lost at least once, and everyone who has lost a bet at least regrets *that* bet in at least the mild sense that if they could have the decision to make it back, knowing the outcome, they’d decline it. In that case ‘problem gambling’ would be made almost synonymous with ‘gambling’ and whatever facts we later said we’d found about problem gambling would really just be facts about gambling in general. On the other hand, suppose we adopted a very stringent definition. For example, we might say that someone only has a clear, serious gambling problem if it leads them to bankruptcy.

We know that if we use that definition, we’re very likely to “discover” that only a tiny proportion of gamblers have problems, so the costs of problem gambling aren’t likely to come out as large by comparison with the benefits derived from less regulated gambling. In both the cases of the overly liberal and the overly stringent definitions, subsequent empirical study doesn’t get allowed to do any real work; we will get out of it as conclusions just what we put into it by definition.

This potential problem caused by careless definition will be obvious enough, but the solution to it is far from obvious. What definition of problem gambling *is* the right one? We’re not going to find this written on a rock somewhere; so it looks as if *any* definition we try will be arbitrary, a pure *decision* we make *prior to* doing any investigative work, which will then drive those very findings.

⁶ That is not, of course, equivalent to actually eliminating it. But almost everyone agrees on what public policy should be when we find it.

Scientists partly overcome this problem by distinguishing between definitions – or, speaking more properly, *analyses* – and *operationalizations*. When you operationalize an idea you don't think of yourself as trying to state the truth about its deep nature. Rather, you just take yourself to be stipulating what you'll mean by it for the duration of some enterprise in which you'll use it. The only test of a good operationalization is usefulness: does it divide the phenomena we're studying into two or more piles that are *both* interesting? Notice that if we used either the overly liberal or overly stringent definitions of problem gambling as operationalizations they'd fail this test: the overly liberal operationalization would put almost all gambling in one pile and in the other pile leave only gambling by people who have hardly ever gambled, and the second pile isn't interesting; the overly stringent operationalization would leave in one pile only gambling by people who went bankrupt as a result.

That pile isn't *completely* uninteresting, but it's not nearly as interesting as the pile we really want, which is the one containing all and only the gambling by people who, if asked "Is your life made significantly worse by your unsuccessful struggles to stop or cut down on gambling?" would answer "Yes".

One way to increase the probability of useful operationalizations is to deploy more than one of them, so as to divide our phenomena into more than just two piles. In the case of patterns in gambling behavior, this is what psychologists have done. Recent common practice among researchers is to attempt to operationalize three terms:

Pathological gambling: A chronic inability to refrain from gambling to an extent that causes serious disruption to core life aspects such as career, health and family. The diagnostic criterion established by the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) is that a person is a pathological gambler if they agree with five or more of the following statements:

1. You have often gambled longer than you had planned.
2. You have often gambled until your last dollar was gone.
3. Thoughts of gambling have caused you to lose sleep.
4. You have used your income or savings to gamble while letting bills go unpaid.
5. You have made repeated, unsuccessful attempts to stop gambling.
6. You have broken the law or considered breaking the law to finance your gambling.
7. You have borrowed money to finance your gambling.
8. You have felt depressed or suicidal because of your gambling losses.
9. You have been remorseful after gambling.
10. You have gambled to get money to meet your financial obligations.

Problem gambling: Researchers want this idea to capture people whose gambling behavior is at least a *nuisance* to them, and is so along the same dimensions as are used to operationalize pathological gambling. We want these operationalizations to line up in this way so that we can use the concept of problem gambling to clearly pick out people who are at risk of becoming pathological gamblers. US President Clinton's Committee on the Social and Economic Impact of Pathological Gambling⁷ operationalized problem gambling as "gambling behavior that results in any harmful effects to the gambler, his or her family, significant others, friends, co-workers etc." In this they simply reflected what they found in the research and treatment literature.

But the operationalization is too liberal to be truly useful; losing R5 on a football bet or being late for lunch with a co-worker because of a queue at the betting window constitute 'harmful effects', however trivial. It would be natural, and maximally helpful, if we could operationalize problem gambling as 'pathological gambling lite' (which might indicate incipient pathological gambling) by describing the problem gambler as someone who agrees with some number of the statements in the operationalization of pathological gambling less than five. However, this confronts us with the problem that the statements don't all seem to be equally diagnostic. Anyone who agrees with (5), (6) or (10) probably has or had a relatively serious problem associated with gambling, and likewise for (8) if 'depressed' is interpreted clinically. But this cannot be said of the other statements unless their interpretations are specially restricted. It cannot be said of (1) under *any* plausible restriction.

Disordered gambling: This denotes the inclusive disjunction of the two ideas above, i.e., gambling that is either pathological gambling or problem gambling.

We said that researchers "attempt to" operationalize these three terms, and it will now be evident why we used this cautious formulation. Pathological gambling is indeed operationalized by DSM-IV. But problem gambling is not; nor can it readily be operationalized by reference to the operationalization of pathological gambling. And then the idea of disordered gambling inherits all the problems attaching to the lack of clarity around problem gambling.

The lack of a settled operationalization for any concept other than pathological gambling raises a problem for the writing of a survey like the present one. The point of a survey is to try to consolidate and compare many studies. This is precisely the activity in which one most depends on confidence that different groups of researchers mean exactly the same thing by every term they use to describe the phenomena they take themselves to be measuring.

We could get around this problem, in principle, by strictly confining our discussion to pathological gambling, remaining absolutely silent on anything else. This, however, would be deeply unsatisfactory procedure from the point of view of readers concerned with policy. Suppose, as seems likely, that there are substantially more problem gamblers, on some reasonable but necessarily (at this point) vague understanding of that concept, than there are pathological gamblers. In that case, the lion's share of the social costs associated with a given gambling policy might be accounted for by the problem gamblers, and no major policy recommendations could be based on studies that ignored them. That would be a situation we'd just have to live with for now if it could honestly be said that the study of problem gambling hasn't yet progressed far enough to license any policy recommendations. But we don't think that *can* honestly be said.

⁷ This Committee was charged in 1996 and reported in 1999. Its report is published as *Pathological Gambling: A Critical Review*. Washington: National Academy Press.

Fortunately, we believe that the particular disciplinary perspective from which this survey is written, that of behavioral economics (henceforth 'BE', to be described in the next section) affords a way around the difficulty. We do not claim it is a complete or all-purpose solution, but we think it is good enough to let us get on with business. In section 1.6 below we will introduce the BE-inspired concept of a 'personal rule'. We'll operationalize 'problem gambling' in terms of this idea. We will use 'pathological gambling' as per the DSM-IV operationalization. Since we'll thus have operationalizations of both 'problem gambling' and 'pathological gambling' we'll be able to use 'disordered gambling' in the way indicated above.

We explained that an operationalization is not yet an analysis. This is important: whereas an operationalization is a tool for doing scientific study and therefore must precede it, analysis is one of the things one wants to get *from* scientific study. That is, scientific investigation aims at telling us, among other things, what problem gambling and pathological gambling 'really are'. This is most usefully put in terms of policy-relevance.

Policy-makers and industry participants need to know whether there's a qualitative 'jump' between problem gamblers, as just typical consumers of gambling services who sometimes have bad days at the casino – as people have bad days in all their activities – and pathological gamblers as people for whom special policies and practices need to be adopted. (Most carefully regulated and administered casinos *do*, at present, implement an assumption like this as a prudential matter and aim to identify and then exclude pathological gamblers. They don't, obviously, exclude or have special policies for those who merely occasionally gamble a bit more than they intended to.) If there is not such a qualitative jump and all people who gamble are on a smooth continuum without 'bright lines', then policy problems are going to be relatively more challenging. Louise Sharpe points out that we get divergent interpretations of some research data depending on which of these hypotheses we consider most probable.⁸ Addressing this question about continuity is one of the basic underlying aims of our survey.

The perspective we'll defend in this survey with respect to problem, but not pathological, gamblers is that there are probably no systematic personal differences between them and people who like to gamble from time to time but haven't experienced problems. (People who don't like to or for some other reason never gamble probably *do* have distinctive properties, but obviously don't matter to the issues with which we are concerned.) We'll maintain, that is, that problem gambling isn't a property of a kind of person, but of a typical person in a kind of situation. Where pathological gamblers are concerned, we'll maintain, matters are different. Surveys from a number of countries and regions of countries tell us that between 1% and 3% of the population of any country or province – sized jurisdiction will tend to be incapable of gambling nearly as moderately as they themselves wish to unless they're assisted with clinical intervention. In Chapter 3 we'll present reasons for thinking that (i) a substantial part of this population, perhaps the majority, are *addictive* gamblers, in that they have brains that physically malfunction in the same way that drug addicts' brains do; and that (ii) for identified *biological* reasons (and not merely as a matter of statistical regularity) people who satisfy the DSM-IV operationalization of pathological gambling are at direct risk of becoming addictive gamblers. We'll show why we think it will very soon be possible to reliably diagnose people in the addictive group by (non-invasively) examining their brains (or perhaps even their genes), and why there are also grounds for optimism that the day isn't far off when they'll be able to control their problem

⁸ Sharpe, L. (2002). A reformulated cognitive-behavioral model of problem gambling: A biopsychosocial perspective. *Clinical Psychology Review* 22: 1-25.

with medication. We'll suggest that this will make policy decisions relatively straightforward where this group is concerned.

There is a good deal of careful ground to cover before we'll be able to get to this. In advance of arriving at a basis for categorization that is supported by a marriage of theory, systematic observation and experiment, behavioral researchers must gather subjects for study and sort them according to some operationalization or other. For this they use what are called 'screens', questionnaires that are administered to people in order to classify them for data-gathering and experimental purposes.

Here we encounter another problem related to definitions of phenomena. With respect to conditions that cause serious harm, such as disordered gambling, the clinical imperative – the need to help suffering people – has higher social priority than scientific research.⁹ In screening people to determine who are pathological or problem gamblers for clinical purposes, we want to err on the side of avoiding 'false negatives'. That is, we'd rather diagnose risk in some people who aren't really in trouble than miss people who need help. Clinical screens build in this bias, and are criticized if they don't. But then we face a dilemma in choosing a screen for research purposes. If we use the clinical screen, we will find in most populations we survey a higher prevalence of the target condition than is really there.

If we instead develop a research screen that differs from clinical screens, then we'll get mis-matches between clinical and research samples. This is problematic for two reasons. First, subject recruitment for studies of conditions that affect only small proportions of the population, such as pathological gambling, is difficult and expensive. Practical purposes require that we usually rely on samples already assembled (often by self-selection) for treatment purposes (so-called 'convenience samples'). This is especially true, for obvious reasons, when an aspect of a research project is evaluation of a therapeutic approach. Second, there are serious ethical problems in sorting clinical populations into those we scientifically classify as 'really' manifesting the condition for which they are being treated and those we admit we have merely diagnosed for cautionary reasons. Since scientific research on subjects cannot be concealed from those subjects, doing this would send signals to the subjects as patients, and thereby undermine the point of having designed our clinical screens to minimize false negatives in the first place.

These objections carry the day in practice. In the study of disordered gambling, research subjects are generally recruited and pooled using clinical screens. This has several implications:

We should be cautious about estimating prevalence. One can do so only by application of a 'double filter'. That is, if one has run statistical tests to determine the *extent* to which a screen gathers false positives, one can then use this to systematically down-weight prevalence estimates in a population obtained using that screen.¹⁰ Notice that this requires

⁹ Note that giving greater urgency and higher priority claim on scarce resources to research might actually yield higher human benefits overall, once the long run is considered. But this would involve treating present sufferers as means to be sacrificed on the altar of a greater future good, which is incompatible with ethical requirements in circumstances in which the institutions of health research and health service are tightly intertwined.

¹⁰ For examples of this see: Cox, B., Enns, M. & Michaud, V. (2004), Comparisons between the South Oaks Gambling Screen and a DSM-IV-based interview in a community survey of problem gambling, *Canadian Journal of Psychiatry* 49: 258-264; Grant, J. Steinberg, M., Kim,

that we have already made some progress in the direction of analysis; if you have nothing beyond a bare operationalization of your phenomenon with which to work, then you have no distinction on which you can base the statistical testing. To a considerable extent, this is where we still stand with respect to research on disordered gambling. This is why, for example, we had to report prevalence for pathological gambling as ‘between 1% and 3%’ - not a very exact figure, since the upper bound is three times the size of the lower bound. The problem isn’t lack of data; it’s ineliminable error when all one has to work with are operationalizations instead of analyses.

When considering any research report, especially of research into the efficacy of a therapy or policy, one needs to know where the screen used to recruit and pool subjects stands comparatively on the question of the average *severity* of pathology found in these subjects.

Consider two hypothetical screens A and B. Suppose that screen A consistently produces samples in which the median subject agrees with 8 statements in the DSM-IV operationalization of pathological gambling, while the median subject recruited by screen B agrees with 6 of them. If one knows that, then one also has reason to suspect that research using screen B is more likely to find that a given policy or therapy is effective than is research on that same policy or therapy that recruits and pools subjects by means of screen A – because the A-screened samples will contain higher proportions of ‘hard cases’. (One can’t be quite sure of this, however, because of the fact that, as noted above, not all the DSM-IV criteria seem to be of equal diagnostic weight. Suppose, for example, that more of the B-screened subjects agree with statements (5) and (6).¹¹)

One of the ‘big’ questions about disordered gambling we identified above as central at the moment is whether there is a qualitative ‘jump’ – in scientific language, a discontinuity – between sub-groups of disordered gamblers. As we have also explained, if there is such a discontinuity then it is highly unlikely that our current operationalizations exactly line up with it; bringing our operationalizations closer into line with it, so that they approach analyses, is precisely one of our scientific goals. This means that we can’t fully trust results derived from work using just one screen, because each screen will have its own bias with respect to the population proportions it selects on either side of the discontinuity – but we can’t know in advance where these biases are, at least until we’ve clearly identified the basis of the discontinuity. To complicate matters further, one team of researchers has given reasons for thinking that there is *more* than one kind of pathological gambler, thus suggesting the possibility of two scales with discontinuities at different measurement points and of different magnitudes.¹² We should therefore ideally run every study we think is important on several groups of subjects recruited using different screens.

S.W., Rounsaville, B. & Potenza, M. (2004), Preliminary validity and reliability testing of a structured clinical interview for pathological gambling, *Psychiatry Research* 128: 79-88.

¹¹ Some research suggests that this would in itself show us that the hypothetical B-screen was defective. A team from the University of Chicago compared responses to each DSM-IV criterion with overall scores on a DSM-IV derived screen, using samples of randomly selected people and people approached on their way to a casino. In both samples, a positive response on criterion (6) was an overwhelmingly reliable predictor of a screen score indicating pathological gambling. Unfortunately, the screen used in this study did not directly probe with respect to criterion (5), which to behavioral economists is the most important one. See Toce-Gerstein, M., Gerstein, D. & Volberg, R. eds. (2003), A hierarchy of gambling disorders in the community, *Addiction* 98: 1661-1672.

¹² Blaszczynski, A. & Nower, L. (2002), A pathways model of problem and pathological gambling, *Addiction* 97: 487-499, present reasons for thinking that there are three different

This set of circumstances confronts us with a trade-off. On one hand, it is very helpful for cross-comparability of studies that one particular screen has so far dominated disordered gambling research. On the other hand, this also limits our ability to bootstrap our way around the inadequacies of our initial operationalizations. In an ideal world we'd have a *matrix*: we would use all available screens on sub-sample populations in all studies, thus having cross-comparability *and* controlling for screen biases.

However, we don't live in the ideal world. Here in the actual one, the dominant instrument mentioned above is the South Oaks Gambling Screen (SOGS) introduced in 1987.¹³ The SOGS is closely based on the DSM-III-R¹⁴ criteria for pathological gambling. (It has not been updated to reflect changes introduced with DSM-IV.) As of a 2000 survey, these diagnostic criteria had been used in over 90% of published research on disordered gambling.¹⁵ The SOGS can be self-administered by prospective subjects, which is convenient and cost-efficient. Petry¹⁶ recommends that for research purposes SOGS be used as an initial screen, and that a more discriminating instrument then be employed as a further filter. She reviews four such existing instruments,¹⁷ among which the comparatively new Diagnostic Interview for Gambling Severity (DIGS)¹⁸ seems especially promising. Another recently developed more intensive instrument is the Canadian Problem Gambling Index (CGI),¹⁹ which facilitates distinguishing between pathological and problem gamblers. Its psychometric properties as measured to date are encouraging.²⁰

We want to make it clear, however, that such double-filtering hasn't thus far been the norm in disordered gambling research, and that establishing it as the norm would add significantly to the average cost of studies. Instruments like the DIGS and the CGI are time-consuming to administer. Their use thus multiplies research staff time, laboratory

kinds of pathological gambler, where each kind results from a different psychological process and starting point, and where people suffering from each kind respond best to different sorts of intervention. We'll engage with this thesis at various points in the report.

- ¹³ See Lesieur, H.R. & Blume, S.B. (1987), The South Oaks Gambling Screen (The SOGS): A new instrument for the identification of problem gamblers, *American Journal of Psychiatry* 144: 1184-1188.
- ¹⁴ This was the edition of the *Diagnostic and Statistical Manual* that was intermediate between the third and fourth editions.
- ¹⁵ Dickerson, E. and Baron, E. (2000). Contemporary issues and future directions for research into pathological gambling. *Addiction* 95: 1145-1159.
- ¹⁶ Petry, N. (2005). *Pathological Gambling: Etiology, Comorbidity and Treatment*. Washington: American Psychological Association, p. 39.
- ¹⁷ Ibid, pp. 39-44.
- ¹⁸ Winters, K.C., Specker, S.M. & Stinchfield, R.D. (2002), Measuring pathological gambling with the Diagnostic Interview for Gambling Severity (DIGS). In J.J. Marotta, J.A. Cornelius and W.R. Eadington, eds., *The Down-side: Problem and Pathological Gambling*, pp. 143-148. Reno: Institute for the Study of Gambling and Commercial Gaming, University of Nevada.
- ¹⁹ J. Ferris and H. Wynne (2001). The Canadian Problem Gambling Index Draft User Manual. www.ccsa.ca/pdf/ccsa-009381-2001.pdf.
- ²⁰ J. Ferris and H. Wynne (2001). The Canadian Problem Gambling Index Final Report. www.ccsa.ca/pdf/ccsa-008805-2001.pdf.

usage time and the amount of subjects' time that must be requested. It thus also multiplies all of the main items in the cost budget of a typical behavioral study.

In many cases even double-filtering of the sort just discussed is inadequate. A number of studies have aimed at identifying the extent to which pathological gamblers tend to have additional psychological / behavioral problems ('comorbidities'), such as substance addiction,²¹ depression,²² schizophrenia,²³ and anti-social personality disorder.²⁴ Determination of comorbidity rates that are stable across populations of pathological gamblers, if there are any such rates, would be helpful for a number of reasons. First, it would assist in predicting the likelihood that a given person is at risk of becoming a pathological gambler, since establishing the comorbidity factors in question would carry at least some information about this. Second, it would be highly relevant to design of treatments and interventions, possibly helping to explain patterns of success and failure. (For example, interventions that work for pathological gamblers who lack certain specific comorbidities might fail for others who do.²⁵) Third, efforts to explain such stable comorbidity patterns as we find might lead us to the explanation of pathological or addictive gambling itself, to the extent that pathological or addictive gambling and comorbidities are sometimes or often consequences of a common causal factor. Comorbidity data are what mainly lead Blaszczynski and Nower²⁶ to the view that there are three different kinds of pathological gamblers. They report that across studies with larger numbers of subjects, stable proportions of pathological gamblers show comorbidity with, respectively (i) nothing, (ii) depression and other mood disorders, and (iii) anti-social personality disorder. Unsurprisingly perhaps, subjects in these groups respond differentially to therapy. They then argue, more controversially, that the different comorbid factors causally contribute to pathological gambling in different ways (hence explaining the varying susceptibility to different interventions).

Obviously, the more we discover about the significance of comorbidity for predicting, explaining and treating pathological and addictive gambling, the more important it will be to screen research subjects for the identified correlated conditions.

-
- ²¹ Crockford, D. & el-Guebaly, N. (1998). Psychiatric comorbidity in pathological gambling. A critical review. *Canadian Journal of Psychiatry* 43: 43-50; W. Feigelman, P. Kleinman, H. Lesieur, R. Millman and M. Lesser (1995). Pathological gambling among methadone patients. *Drug and Alcohol Dependence* 39: 75-81.
- ²² Cunningham-Williams, R., Cottler, L., Compton, W. & Spitznagel, E. (1998). Taking chances – problem gambling and mental health disorders – Results from the St. Louis Epidemiological Catchment Area (ECA) study. *American Journal of Public Health* 88: 1093-1096; Dannon, P., Lowengrub, K., Sasson, M., Shalgi, B., Tuson, L., Saphir, Y. & Kotler, M. (2004). Comorbid psychiatric diagnoses in kleptomania and pathological gambling: a preliminary comparison study. *European Psychiatry* 19: 299-302.
- ²³ Ibid.
- ²⁴ Lesieur, H. (1987). Gambling, pathological gambling and crime. In T. Galski, ed., *The Handbook of Pathological Gambling*, 89-110. Springfield: Charles H. Thomas.
- ²⁵ Ladd, G. & Petry, N. (2003). A comparison of pathological gamblers with and without substance abuse treatment histories, *Experimental and Clinical Psychopharmacology* 11: 202-209, report that pathological gambling is more persistent in subjects who have been treated for substance abuse than in otherwise comparable subjects who have not been so treated.
- ²⁶ Blaszczynski, A. & Nower, L. (2002). A pathways model of problem and pathological gambling. *Addiction* 97: 487-499.

This does not imply anything especially onerous for research methodology where substance abuse is the possible factor in question, since it can be effectively screened for by comparatively basic instruments. However, reliable screening for psychiatric disorders requires use of sophisticated instruments and qualified expert administration and analysis. This would have major implications for the cost and complexity of research.

Note that an approach that focuses on searching for comorbidity factors is almost inevitable to the extent that pathological gambling is conceptualized as a kind of *disease* – that is, as a specific breakdown in the functional constitution of the person. Thus the reason Blaszczynski and Nower conclude from three comorbidity patterns that there are three different kinds of pathological gambling is that they implicitly *identify* any case of pathological gambling with its molecular etiological pathway.²⁷ This assumption sets researchers off in search of the site of damage. They often suppose it to be unlikely that something could go wrong with a person's mind or brain that would produce *only* pathological gambling, in which case such searches are bound to be, at least in part, searches for comorbidity factors. This has tended to be the perspective of much recent psychiatry. Its promised implications for policy and intervention are straightforward. If we could identify the site of damage in the pathological gambler, then we could either seek to repair the damage in question or, if that is not possible, use diagnosis of the damage as a basis for knowing whose access to legal gambling opportunities should be restricted or blocked. And perhaps we could set up mechanisms to identify those people who are at risk of becoming pathological gamblers and effectively intervene to stop onset. In Chapter 3 we'll present evidence that this approach looks like it going to bear fruit where addictive gamblers are concerned.

Where policy is concerned, however, what if there are pathological gamblers who *aren't* addicted? Or suppose we found a reliable diagnostic but not a cure for addictive gamblers? We could probably prevent many of them from going to casinos or racetracks. Preventing them from buying lottery tickets or playing Internet poker would be more difficult. But would it even be estate markets? Should pathological or addictive gamblers be prevented from re-selling houses or other assets?²⁸

This question helps to remind us that gambling *per se* is a normal, not a deviant, kind of human behavior. By this we don't just mean that it is widespread; we mean that a disposition to gamble is adaptive and a disposition to *never* gamble is severely *maladaptive*. Few people aspire to have the kind of personality that tries to avoid all risks. More to the point, few people aspire to not *enjoy* taking some risks. Aversion to risk, or mere inability to appreciate the pleasure of moderate risks, amounts to aversion to novelty and absence of both curiosity and ambition. This suggests that, except where addictive gamblers are concerned, it may be mistaken to take for granted that our main question must be: What, specifically, is wrong with pathological gamblers?

²⁷ This bit of special scientific terminology is hard to avoid, because it has no everyday-language equivalent that isn't quite cumbersome. An 'etiological pathway' refers to a *sequence* of events or conditions, taken to be causally related by a mechanism at each stage of the sequence, that has the state we want to explain as its end-state. 'Molecular' refers to the idea that each element in the sequence is a discrete event, rather than a statistical description of a distribution. We will discuss the significance of assuming conditions like disordered gambling to have molecular explanations in the next section.

²⁸ These are issues with which organizers of Gamblers Anonymous and other treatment organizations wrestle. There is not a consensus on them in the clinical and helping communities.

Is it possible that gambling problems arise, at least in the first place, when people fail to learn to effectively manage some natural dispositions shared by almost everyone? There is a discipline, closely entangled and increasingly integrated with traditional psychology but importantly distinct from it, which begins from this alternative opening perspective. This discipline is called 'behavioral economics'. In our opinion it has for some time now been the most fecund source of insights and productive research procedures for studying the broad class of behavioral dispositions popularly associated as 'compulsions' and 'bad habits' specifically including (at least) non-addictive disordered gambling.

Where addictive gambling is concerned, it has been thought by some of its founders to be in tension with neuroscience, a matter to which we will return – in application to gambling – in section 1.7. Our view is that this tension, however natural, must be overcome, since both BE and neuroscience offer crucial insights into the phenomena of interest to us here, and many others. But before we get to all that, we must digress from our direct focus on gambling to explain to the reader what BE is in the first place.

1.2. Behavioral economics

One of the respects in which pathological gambling (in general) has often been grouped with stereotypical substance dependencies such as alcoholism is by reference to symptoms, including sensed loss of control and preoccupation in consciousness. But they have also been associated by appeal to the way in which these behavioral patterns both supposedly manifest *impulsiveness*. The report of President Clinton's Committee on the Social and Economic Impact of Pathological Gambling says that "there is considerable consensus that gambling involves impulsiveness."²⁹ However, the report is ambiguous with respect to the interpretation of this idea. Discussion of this ambiguity is a useful way of introducing the distinction between traditional psychology as a conceptual framework and the framework of behavioral economics.

According to DSM-IV, an impulse control disorder is manifest in a person to the extent that that person shows recurrent and significant "failure to resist an impulse, drive or temptation to perform an act that is harmful to the person or to others."³⁰ On this interpretation impulsiveness is an 'inner-referenced' pathology, a tendency of some sort of behavioral 'organ' to misfire. Again, this is the perspective that the great psychologist

Richard Herrnstein dubbed 'molecular'. Presupposed here by way of contrast is a notion of 'normally non-impulsive molecular behavioral causation' and it is by implicit reference to this that we are to understand the idea of the subject's doing harm to herself. Traditional psychology often asks us to contrast 'healthy' with 'unhealthy' behaviour in this way.

²⁹ Committee on the Social and Economic Impact of Pathological Gambling (1999). *Pathological Gambling: A Critical Review*. Washington: National Academy Press. p. 37.

³⁰ American Psychiatric Association (1994). *Diagnostic and Statistical Manual IV*. Washington: American Psychiatric Association. p. 609.

The definition just given more or less *identifies* impulsiveness with unhealthy behaviour.³¹ This is in tension with the fact noted at the end of the preceding section that almost everyone gambles, and that a person who absolutely refused to gamble would not likely be regarded as 'healthy'. This implies that 'health' must be compatible with harm caused by gambles, since the idea of a gamble that is guaranteed not to produce harmful consequences makes little sense. Reflecting the different usages found in the literature,³² the Committee's report *also* understands impulsiveness in a different sense: as a tendency to choose smaller earlier rewards over larger later ones. The alcoholic could have sobriety, accomplishment and social respect today and tomorrow, with all the possible streams of subsequent reward these can produce; but she chooses instead the vivid but transient pleasure of a high now, from which she must subtract the costs of hangover and social disapprobrium tomorrow. The disordered gambler could have money to pay his bills, and so secure his welfare for the month, but instead chooses to experience the satisfaction of a brief thrill, from which he must subtract the value of his expected loss.

Describing these scenarios in terms of *choices* immediately introduces into the picture the possible relevance of *economic* analysis. Economics is, in its broadest sense, the study of the way in which people (and other organisms and groups) try to promote their welfare in the face of scarcity. 'Scarcity' has two complementary connotations. First, an economic agent's wants exceed what the environment furnishes, especially given other agents with wants of their own who share that environment. Second, acting so as to capture one stream of benefits typically rules out capturing another possible stream. The value of what is foregone as a result of choice is referred to as the 'opportunity cost' of the choice in question.

Thus conceiving of impulsiveness in the way suggested immediately above amounts to seeing the impulsive person as being prepared to forego larger, later rewards as the opportunity cost of smaller, earlier ones. As soon as we frame things this way, we can introduce all of the technical apparatus of economic analysis. If indulging in a drink or taking a bet is a trade-off between benefits and costs, then we can ask questions about how much change would have to be made to the goods on either side of the ledger to produce the opposite decision.

³¹ Of course, in any specific study impulsiveness will be operationalized more precisely, if only implicitly, by use of a specific measurement scale. Our point here is not to cast doubt on any particular empirical studies, but to make a conceptual point.

³² Evendon, J. (1999), Varieties of impulsivity, *Psychopharmacology* 146: 348-361, records a range of divergent interpretations of the idea found in psychology, psychiatry and ethology, and reports evidence that "several neurochemical mechanisms can influence impulsivity, and that impulsive behaviour has no unique neurobiological basis". We find it hard to tell from Evendon's discussion whether he thinks that all the behavioral aspects that have been associated with impulsiveness are correlated expressions of one phenomenon that has a multifactorial causal profile, or whether he thinks there are several different kinds of impulsiveness that aren't systematically related. He says things that variously indicate both of these incompatible interpretations.

A very large quantity of empirical evidence shows that the behaviour of people popularly thought of as ‘addicts’ is indeed systematically sensitive to relative opportunity costs.³³ In section 1.8, we will return to and explain the significance of this for the way in which we conceptualize the family of disorders in which disordered gambling behaviour has usually been included by psychologists, by the clinical community, and by policy-makers. We will then ask what this perspective suggests to be *distinctive* about disordered gambling as a member of this family.

First, however, we must say more to introduce and explain the nature of behavioral economics itself. A reader new to the literature in this area may find herself a bit confused at first because the term is used in two different ways. Sometimes the phrase ‘behavioral economics’ (BE) is used to denote the whole sub-discipline of economics that empirically, often experimentally, studies patterns in people’s responses to shifts in incentive structures and to changes in the schedules and prices of rewards and costs that they face.³⁴ This can be contrasted with more traditional economics as follows. Economic theories and models in the traditional vein are typically interested in what an *ideally* rational agent *would* do in a given incentive situation, given a particular profile of desires (a ‘utility function’) and a budget constraint. In addition, traditional economics typically searches for generalizations that hold good in any set of institutional conditions – in democracies or dictatorships or states of anarchy; in cases where government successfully stabilizes the value of money and in cases where it doesn’t, in cases, for that matter, where there is no money at all but people nevertheless exchange. (We could add here: in cases where there is legal gambling and in cases where there isn’t.) Against this background, BE is often used to refer to any approach that aims at theories and models of what *actual* people do in *particular* institutional settings. This question needs separate study because actual people are not ideally rational agents, and because real institutions cannot be maintained for free; they impose so-called ‘transaction costs’, and they will survive only to the extent that some agents deem themselves to be best off if they pay these costs.

When people encounter this contrast, they often respond by thinking that behavioral economics should *supplant* traditional economics. Shouldn’t we always care more about what actual people in concrete conditions do than about what ideal – but non-existent – agents in completely abstract situations do? (After all, actual people are never in completely abstract situations.) But this way of putting things is misleading. The concepts we use to understand actual people (or actual firms, or actual governments, or actual agents of other kinds) are given their meaning by their definitions in the general theory of ideal rational agency.

³³ For a survey, see Vuchinich, R. (1999), Behavioral economics as a framework for organizing the expanded range of substance abuse interventions (pp. 191-218), in J. Tucker, D. Donovan and G. Marlatt (eds.), *Changing addictive behavior: Bridging clinical and public health strategies*. New York: The Guilford Press.

³⁴ The standard handbook is Kagel, J. & Roth, A., eds. (1995), *The Handbook of Behavioral Economics*. Princeton: Princeton University Press. A more up to date compendium is Camerer, C., Loewenstein, G. & Rabin, M. eds. (2003), *Advances in Behavioral Economics*. Princeton: Princeton University Press. Thaler, R. (1992), *The Winner's Curse*, New York: Free Press, provides a good popular account, but is now dated.

Without this theory, we wouldn't be able to state clear economic hypotheses about actual people in actual circumstances in the first place. In the present survey, however, all of this will be kept in the background; the reader will not encounter any instances of traditional, purely abstract, economic reasoning in what follows.

We have now identified one of the two things that 'BE' sometimes means. To avoid confusion, we will henceforth refer to this as 'experimental / institutional economics' (EIE), reserving the term 'BE' for something more specific, and more immediately relevant to disordered gambling and to the traditional 'core addictions'. Whereas EIE involves no adjustments in standard economic theory, but merely *restricts* it for application to specific scenarios, BE involves a substantive theoretical innovation. The innovation in question is due mainly to the late Richard Herrnstein.

In standard economic theory, agents are modeled as *maximizing* an abstract function referred to as 'utility'. This is a highly technical and widely misunderstood idea, which cannot be fully grasped or explained independently of the mathematics economists use to write their theories and models.³⁵ The crucial *empirical* implication of the idea that will be relevant here is the following. A utility maximizer will choose *consistently* over time, in the sense that she will vary her choice patterns between a given stream of benefits *B* and all possible opportunity costs of *B* (i.e., other benefit streams) *only* on the basis of changes in two conditions: (i) her stock of *B* in proportion to other goods, or (ii) the externally determined price of *B* relative to prices of other goods. Utility maximization restricts the *directions* of these relationships: the maximizing agent will accept higher opportunity costs for *B* the less of it she has, and she'll choose a greater quantity of *B* the lower its price.

The reader who isn't used to economic reasoning is urged not to puzzle too long over these formulations, because all that will matter here is one of their consequences. This is that *if* an agent knows what consumption opportunities her environment is going to provide, she'll allocate her resources over time in such a way that she'll get the highest possible welfare *overall*. This will typically mean that she'll often be giving up a possible temporary welfare increase at a given time for the sake of this general maximization. If the agent is uncertain about future consumption opportunities, this complicates matters technically, but as long as she calculates her probabilities rationally, the main point remains that she'll choose in such a way that her overall welfare will be maximized so long as she doesn't suffer from improbable bad luck. (The following contrast may help here. An uncertain agent who allocates resources as wisely as possible and then encounters bad luck will be *disappointed* but has nothing to *regret*, unlike the agent who contributes to her own misfortune by suboptimal allocations or failure to use information that was available to her.)

In this framework, addiction must be interpreted just as a sign of preferences that happen to make life hard. If we see a person waste away from alcoholism, we must simply suppose that this person considers drunkenness to be worth very high opportunity costs. Like any other utility maximizer, she gets as much welfare overall as is possible for her. That she becomes less and less *happy* over time is because alcohol consumption causes tolerance to increase; the opportunity cost of the same high rises steadily as the history of drinking plays out.

³⁵ For studies of the technical complications in the concept of utility maximization, see Mandler, M. (1999), *Dilemmas in Economic Theory*, Oxford: Oxford University Press; Ross, D. (2005), *Economic Theory and Cognitive Science: Microexplanation*, Cambridge MA: MIT Press.

So far as the alcoholic's harm to herself is concerned, the only policy response by others that isn't cruel and oppressive is to try to avoid unnecessary increases in the price of drinks, and helping her minimize her opportunity costs by assisting her to avoid situations that are potentially dangerous or embarrassing for drunk people. It would make no moral sense in this framework for others to help the agent avoid alcohol since (once potential harm to second and third parties is minimized) this would simply amount to lowering her welfare. All these implications of the traditional economic model of utility maximization to addiction were originally worked out in technical form by the Nobel Prize winning economist Gary Becker and his collaborators.³⁶

No one can *prove* that people don't maximize so-called 'expected' utility in this way. Expected-utility maximization is a mathematical modeling technology, not a psychological hypothesis, and we should evaluate it like any other tool: can we do the job we want it for – explaining and predicting behaviour – with no more sweat and strain than we'd face if we used an alternative and actually available tool? So-called addiction is one of the leading behavioral patterns – perhaps *the* leading behavioral pattern – for which the answer to this question is 'no'. The problem is that most addicts don't just consume their substance of abuse until they've run out of all resources, as the straightforward application of standard economic theory with expected-utility maximization predicts that they should. Instead, most of them spend resources trying to *stop* consuming their substance of abuse. It isn't *impossible* for economists like Becker to explain this. For example, addicts might want to periodically go dry so they can re-calibrate their tolerance levels and then enjoy their highs less expensively for awhile. The real problem for expected-utility maximizing accounts of addiction is that *most addicts successfully quit eventually*. Becker-style theories *must* infer from this that addicts did not initially know that addiction was going to be bad, and only learned this through experiencing lower welfare levels than they expected. Again, this is possible. But now it's relapse that becomes puzzling. We have to suppose that each time an alcoholic stops drinking but then relapses she's aiming to get drunk more cheaply for awhile, but then, when she at last stops once and for all, this was because she *finally* realized that she was doing herself more harm than good. If she never *does* stop once and for all we have to suppose that she was *not* doing herself more harm than good, in her own terms.

To summarize all this: on the traditional economic accounts, we can *either* tell the story in a way that makes reasonable sense of addicts' tendencies to relapse but not of the fact that most eventually recover; or vice versa. What we cannot do is explain *both* of these phenomena in a way that doesn't involve completely *ad hoc* psychology for which there is absolutely no evidence.

³⁶ See Becker, G. (1976), *The Economic Approach to Human Behavior*, Chicago: University of Chicago Press; Becker, G. and Murphy, K. (1988), A theory of rational addiction, *Journal of Political Economy* 96: 675-700.

Almost all people with experience of addicts and addiction find this to be too much sweat and strain to make the traditional tool of expected-utility maximization worth using, especially because Herrnstein provided us with an alternative one.³⁷ He spent his career showing that we get accurate predictions of most human and animal choice behavior if we instead model them as maximizing *local* rather than *overall expected* utility. He called the pattern of behavior we observe on this model *melioration*. The family of models that technically represent melioration share commitment to what Herrnstein dubbed the ‘matching law’. This says that if a person or animal can distribute their resources to a range of activities over what they account as an interval of time, they’ll do so in exact proportion to the magnitudes of rewards they experience for each activity *in that time interval*. BE, in the specific sense for which we will use that label here, is the intersection of economics and psychology that models behavior by applications of the matching law.

Herrnstein did not merely show that this way of modeling *addicts’* behavior captures the data; he demonstrated that it applies to people and other animals *in general*. People popularly thought of as addicts are merely ideally vivid cases for illustration. All agents who match and meliorate will tend to behave, from the traditional economist’s point of view, short-sightedly, failing to take full account of their long-run welfare.

We must say ‘tend to’ rather than the more straightforward ‘will’ for two reasons. First, the matching law and expected-utility maximization will predict the same behaviour whenever we focus on an agent choosing between two streams of benefits *A* and *B* in which *A* is more rewarding than *B* over *all* intervals. For example, an alcoholic will display consistent choice as between good whiskey and bad whiskey. Second, many human institutions are set up so as to incentivize people who meliorate to act as if they were maximizers; financial institutions are the outstanding examples here, finding devices like interest payments and derivatives to ensure that all failures to respect the logic of maximization can be efficiently exploited for someone else’s profit. (This simultaneously ensures that resources don’t go to waste and that investors have incentives to adjust their accounting of time intervals until they’re acting like maximizers.) But for the phenomena of concern to us here – first, addiction-like behaviour patterns in general, and then disordered gambling specifically – BE makes distinctive predictions and offers distinctive explanations. As we will see, these predictions and explanations have been very strongly supported by empirical behavioral research.

Most importantly, BE explains behavioral patterns in a so-called *molar*, rather than a *molecular*, way. It doesn’t aim to trace out causal sequences whereby some particular event in a person’s life, or in their brain, or in their consciousness, leads to this particular drink or to that particular refusal of a drink, to this specific ante at the blackjack table or that specific decision to go to a movie tonight instead of to the casino. What BE explanations aim to account for are *patterns* of behavior, the *relative* opportunity costs that people pay over a given stretch of time for different rewards. Pathological gamblers are, in the first instance, people who pay what seem to the rest of us to be – and what they typically eventually agree themselves to be – extravagant opportunity costs for whatever it is they find rewarding about gambling.

³⁷ The most important papers of Herrnstein and his immediate collaborators have been posthumously gathered together as a book: Herrnstein, R. (1997), *The Matching Law*, Cambridge MA: Harvard University Press.

They don't offer this payment on every single possible occasion, but that isn't what BE aims to explain. What is of interest in BE is that they offer the payment significantly more often than problem gamblers, who in turn offer it significantly more often than non-disordered gamblers. Note that on this perspective it makes no sense to think of the returns to gambling – either the excitement or the winnings, whichever turn out to matter more motivationally – as *bad* things for the gambler. They are, after all, what the gambler *likes* – what she gambles *for*. Indeed, almost everyone gambles for these rewards to *some* extent, regardless of whether they ever play structured games of chance and pay third parties to set them up. What is bad, for the gamblers, about disordered gambling is that they pay more for these rewards, over time, than they themselves want to.

BE is a *general* framework for thinking about behaviour. As will be discussed in the next few sections, melioration and matching are compatible with several specific models of the phenomena traditionally grouped together as 'addictions', including pathological gambling. It also implies a particular set of experimental *methods* for trying to choose among these different models. We describe these experimental approaches, and what has been learned about disordered gambling from their application, in Chapter 2.

1.3. Behavioral economics of substance use and addiction

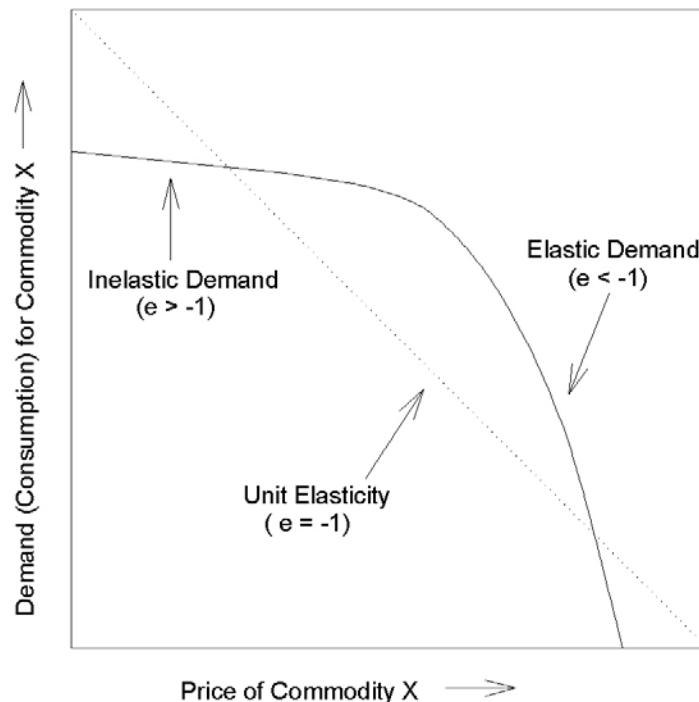
The scientific community has not reached a consensus regarding whether or not disordered gambling and substance abuse are fundamentally the same type or are fundamentally different types of behavioral disorders. (We will give reasons for thinking that such a consensus is now becoming more likely in Chapter 3.) To the extent they are similar, then knowledge gained regarding substance abuse would also apply to gambling.

To date, BE has been much more extensively applied to substance abuse than to gambling, so we will summarize some of the BE work on substance abuse as an illustration of what can be gained from the BE perspective. Because BE is a *choice* framework, the value of substance use, and therefore the extent to which it is chosen, is viewed as being determined by the benefits/costs of substance consumption in relation to the benefits/costs of other available activities. This research on substance abuse has included approaches based on EIE as well as on BE innovations derived from the matching law. We will summarize the former first, which has primarily consisted of investigations of the effects on substance consumption of (1) direct constraints on access to the substance and (2) constraints on access to other valuable activities.

Effects of constraints on access to the substance. The vast majority of this research has focused on the prices of substances that people abuse, and has employed a research strategy known as "demand curve analysis." Demand is the amount of a commodity that is purchased, and a demand curve displays how purchases of a commodity change as its price changes. Figure 1 shows a hypothetical demand curve. In Figure 1, the price increases from left to right along the x-axis, and demand increases from bottom to top along the y-axis. The *law of demand* states that there will be an inverse relation between price and demand, such that higher prices lead to lower demand (and vice versa). As is typical of EIE, precise quantitative concepts and methods have been developed to study these relations. One such concept is *elasticity of demand*, which is a quantitative description of the degree of changes in demand in the face of price changes. Elasticity in this context is a metaphor for how much "curve" or "bend" exists in the demand curve, with inelastic curves having little or no bend and elastic curves having much bend.

In Figure 1, the dotted line labeled unit elasticity goes straight from the upper left to the lower right of the graph. This indicates that demand is reduced in exact proportion to any price increase. But most demand curves are more like the curved line in Figure 1, with demand changing very little (inelastic demand) at the low end of the price range but changing very much (elastic demand) at the high end of the price range.

Figure 1
Own-Price Elasticity of Demand



Prior to the work on demand for substances of abuse, it was not known if such meant in regarding these substances as “addictive” was that the demand law didn’t apply to them, and so consumption would continue at high rates regardless of price increases. That turned out not to be the case, and this research on substance use has demonstrated the value of such demand curve analyses. Bickel and colleagues³⁸ re-analyzed data from numerous drug consumption experiments (including cocaine, codeine, d-amphetamine, ethanol, ketamine, methohexital, morphine, pentobarbital, phencyclidine, and procaine) by using demand curves. They found that the demand curves revealed the typical mixed elasticity for all drugs studied, with demand being inelastic at lower prices and elastic at higher prices (like the curved line in Figure 1).

³⁸ Bickel, W.K., DeGrandpre, R.J., & Higgins, S.T. (1993). Behavioral economics: A novel experimental approach to the study of drug dependence. *Drug and Alcohol Dependence* 33: 173-192; Bickel, W.K., DeGrandpre, R.J., Higgins, S.T., & Hughes, J.R. (1990). Behavioral economics of drug self-administration: I. Functional equivalence of response requirement and drug dose. *Life Sciences* 47: 1501-1510; Bickel, W.K., Madden, G.J., & Petry, N.M. (1998). The price of change: The behavioral economics of drug dependence. *Behavior Therapy* 29: 545-565.

Similar quantitative relations were found when data from 17 studies of human cigarette smoking were reanalyzed³⁹, and in an experiment with human cigarette smokers that explicitly employed a demand curve analysis⁴⁰. This same inverse relation between substance consumption and price has also been observed in the natural environment for alcoholic beverages⁴¹, cigarettes⁴², and illicit drugs⁴³.

In summary, this research on substance use has consistently found a quantifiable relation between consumption and price such that consumption decreases as price increases (and vice versa), and this relation has consistently been found in different species, for different abused substances, in normal and clinical (“addicted”) populations, and in laboratory and natural environments. It is particularly important to note that the relations between consumption and price also hold for individuals “addicted” to the substances. Such findings indicate that the consumption of abused substances can be usefully described with the same analytic tools that apply to all commodities.

Policy note. The quantification provided by drug demand curves in laboratory experiments has provided valuable information for assessing abuse liability (the extent to which a drug may become the object of excessive consumption) and in developing pharmacological interventions for drug abuse.

For example, using a demand curve analysis, Bickel and colleagues⁴⁴ showed that demand for pentobarbital was over 12 times more responsive to price increases than was cocaine, which means that for every unit of price increase, demand for pentobarbital was reduced 12 times more than the demand for cocaine. Thus, demand for pentobarbital was found to be much more elastic than demand for cocaine, which implies that pentobarbital has less abuse potential than cocaine. Of course, this finding matches casual observation regarding the abuse potential of the two substances. Given that different games of chance also vary in their potential to generate disordered gambling, demand curve studies that compared demand for gambling opportunities across games could yield valuable

-
- ³⁹ DeGrandpre, R.J., Bickel, W.K., Hughes, J.R. & Higgins, S.T. (1992). Behavioral economics of drug self-administration: III. A reanalysis of the nicotine regulation hypothesis. *Psychopharmacology* 108: 1-10.
- ⁴⁰ Bickel, W.K., DeGrandpre, R.J., Hughes, J.R. & Higgins, S.T. (1991). Behavioral economics of drug self-administration: II. A unit-price analysis of cigarette smoking. *Journal of the Experimental Analysis of Behavior* 55: 145-154.
- ⁴¹ Leung, S.-F. & Phelps, C.E. (1993). “My kingdom for a drink . . . ?” A review of the estimates of the price sensitivity of demand for alcoholic beverages. In M.E. Hilton & G. Bloss (Eds.), *Economics and the Prevention of Alcohol-Related Problems* (pp.1-31). Rockville, MD: National Institute on Alcohol Abuse and Alcoholism Research Monograph No. 25. NIH Pub. No. 93-3513.
- ⁴² Chaloupka, F.J. (1991). Rational addictive behavior and cigarette smoking. *Journal of Political Economy* 99: 722-742.
- ⁴³ Chaloupka, F.J., Grossman, M. & Tauras, J.A. (1999). The demand for cocaine and marijuana by youth. In F.J. Chaloupka, W.K Bickel, M. Grossman, & H. Saffer (Eds.), *The Economic Analysis of Substance Use and Abuse: An Integration of Econometric and Behavioral Economic Perspectives* (pp. 133-156). Chicago: University of Chicago Press.
- ⁴⁴ Bickel, W.K., DeGrandpre, R.J.. & Higgins, S.T. (1993). Behavioral economics: A novel experimental approach to the study of drug dependence. *Drug and Alcohol Dependence* 33: 173-192.

information regarding the games' "abuse liability." This is work that could usefully be commissioned from behavioral economists by industry participants or regulators.

Effects of constraints on access to alternative activities. This choice perspective implies that the relative value of substance use will also depend on what other activities are available to potential abusers and on the benefits/costs of those activities. Understanding the contribution of this more general context is a key issue in substance use research, because strong preferences for substances arise in natural environments that presumably contain opportunities to engage in a variety of other activities. In general, substance use should be determined in part by the benefits/costs of other activity opportunities, with substance use increasing as the value of other activities decreases. A number of laboratory studies with both animal and human subjects have demonstrated this relationship with a variety of substances.⁴⁵

The same relation found in the laboratory between substance abuse and alternative activities should also be found in the natural environment. Vuchinich and Tucker⁴⁶ applied this reasoning to an analysis of relapse episodes in a clinical sample of alcohol abusers after treatment. They found that relapses occurred more often when the rewards (e.g., intimate, family, social relations) in the recovering individuals' environments changed from more to fewer alternative rewards being available. Also, decreasing substance abuse and maintaining recovery can be thought of as involving the reverse process wherein the reward structure shifts from fewer to more alternative rewards being available. Such relationships between the maintenance of recovery and increased access to valued rewards other than drinking have been found in community samples of alcohol abusers⁴⁷ and in heavy-drinking college students⁴⁸

Policy note. A simple yet potentially powerful implication of the BE perspective for clinical interventions is that more attention should be paid to the availability of valuable nonsubstance-related activities and the extent to which access to them depends on abstinence or reduced substance use.⁴⁹ The research summarized above clearly shows that the variety and availability of valuable activities other than substance use can have a critical influence on substance consumption, and these relations have important relevance for the development of clinical assessment and intervention procedures and policies.

-
- ⁴⁵ For a review see Carroll, M.E. (1996). Reducing drug abuse by enriching the environment with alternative non-drug reinforcers. In L. Green & J. Kagel (Eds.), *Advances in Behavioral Economics: Vol. 3. Substance Use and Abuse* (pp. 37-68). Norwood, NJ: Ablex.
- ⁴⁶ Vuchinich, R.E., and Tucker, J.A. (1996). Alcoholic relapse, life events, and behavioral theories of choice: A prospective analysis. *Experimental and Clinical Psychopharmacology* 4: 19-28.
- ⁴⁷ Vuchinich, R.E. & Tucker, J.A. (1996). Life events, alcoholic relapse, and behavioral theories of choice: A prospective analysis. *Experimental and Clinical Psychopharmacology* 4: 19-28.
- ⁴⁸ Murphy, J.G., Correia, .C.J., Colby, S.M. & Vuchinich, R.E. (2005). Using behavioral theories of choice to predict drinking outcomes following a brief intervention. *Experimental and Clinical Psychopharmacology* 13: 93-101.
- ⁴⁹ For a review see: Higgins, S.T., Heil, S.H. & Lussier, J.P. (2004). Clinical implications of reinforcement as a determinant of substance use disorders. *Annual Reviews of Psychology* 55: 431-461.

It also is quite possible that similar relations exist between changes in disordered gambling and alternative activities, and that more attention should be paid to these relations in clinical interventions for gamblers. It would be interesting to have, for example, household surveys and subsequent microeconomic analyses that tested the effect of new sports facilities or entertainment complexes in communities where there were already organized gambling venues. Note that these effects would probably vary from one culture or country to another.

Summary. This very conventional BE work on substance use and abuse has been important because it demonstrated that abused substances obey the same economic laws as other, “nonaddicting” substances. That is, even in “addicts,” consumption goes down as price goes up (and vice versa), and consumption goes up if valuable alternative activities become more difficult to obtain (and vice versa). This in turn establishes that BE provides an appropriate framework for the study of addiction: addicts are economic agents in the same sense as non-addicts, even though there are important differences between their patterns of choice.

1.4. The matching law and temporal discounting

The outcomes of our choices sometimes occur immediately, but more often the outcomes do not occur right away. Most often, we choose to engage in activities now that produce positive and/or negative outcomes sooner or later in the future. The value of delayed outcomes is an especially important issue for understanding substance use and addiction, and also gambling, because individuals are repeatedly choosing between an immediately available but relatively small reward (substance consumption or the fun of betting) and engaging in alternative activities that will produce more delayed but more valuable rewards in the future (e.g., positive intimate, family, or social relations; vocational or academic success; purchase of something wanted and saved for).

In these types of choices, gaining access to the more valuable but delayed rewards is facilitated by foregoing the less valuable but sooner reward; and, conversely, repeatedly enjoying the smaller sooner rewards risks reduced access later to the larger later reward. When someone’s choices in these life areas go awry (e.g., excessive drinking or gambling at the expense of marriage and family), it is almost irresistible for us to introduce *irrationality* into the analysis. Most such individuals are in fact irrational, in the strict sense of the economist, as defined earlier in this chapter. Like the fanciful cobraholio, their preferences are inconsistent over time; at Time A they (truthfully) report a strong desire to drink less or not at all in order to maintain a happy marriage and family life, but at Time B they drink to excess and thereby further threaten the very things they desired at Time A. Research derived from the matching law has produced tremendous strides in our understanding of such choices and their disorders.

We will now describe the theoretical and empirical basis of this significant progress. In order to do so, we need to delve into some details of the analysis. This gets a little complicated, but the mathematics we’ll introduce here are no more complex than high school arithmetic, and we believe a firm grasp of the basics will reap benefits for the reader in later chapters.

Experiments on the matching law⁵⁰ typically give subjects opportunities to respond on option 1 and option 2, manipulate some aspect of reward (e.g., frequency, amount, or delay) derived from option 1 and option 2 choices, then relate the pattern of choices to the pattern of rewards received. In these experiments, both animals and humans make a very large number of choices across multiple sessions, each of which lasts about one hour, and the individual rewards typically are a small amount of food for animals and a small amount of money for people. The first aspect of reward studied was frequency, with the same size of reward occurring more or less frequently from option 1 and option 2 choices. Those studies found, for example, that if reward occurred twice as frequently from option 1 as from option 2, then option 1 was chosen twice as often as option 2. Because the proportion of choices for each option *matched* the proportion of rewards received from each option, the relation was termed the *matching law*. Later studies investigated varying the amount and varying the delay aspects of reward, and produced results similar to those that studied frequency of reward. For amount, if rewards occurred equally often from option 1 and option 2, but the option 1 rewards were, for example, twice as big, then option 1 was chosen twice as often as option 2. For delay, when the same size rewards occurred equally often on option 1 and option 2, but receipt of the option 2 rewards were, for example, delayed twice as long as receipt of the option 1 rewards, then option 1 was chosen twice as often.

The huge leap forward in our understanding of impulsiveness and self-control occurred when these scientists began studying choice between rewards that varied in both amount and delay, with rewards from one option being smaller but sooner and rewards from the other option being larger but later. Equation 1⁵¹ is a version of the matching law (see section 1.2), and is a very accurate description of the relation between choice and rewards that vary in both amount and delay. In Equation 1, B_1 and B_2 represent the number of choices of option 1 and option 2, A_1 and A_2 represent the amount (size) of each reward received from option 1 and option 2, and D_1 and D_2 represent the delay between the choice and receipt of the reward from option 1 and option 2. This equation shows that relative choice of the two options is directly proportional to the relative amount of reward and inversely proportional to the relative delay of reward.

$$\frac{B_1}{B_2} = \frac{A_1}{A_2} \times \frac{D_2}{D_1} \quad (1)$$

This relation shows that there is an unsurprising tendency in both animals and people to prefer larger rewards to smaller rewards, and to prefer sooner rewards to later rewards. But these two tendencies are put in conflict when the smaller reward is sooner and the larger reward is later. Choice of the smaller, sooner reward (SSR) is labeled impulsive, and choice of the larger later reward (LLR) is labeled as self-controlled⁵². The critical importance of this form of the matching law for understanding impulsiveness and self-control is the manner in which the value of rewards is predicted to vary with their delay.

⁵⁰ For a thorough account see: Rachlin, H. and Laibson, D.I. (1997). *The Matching Law: Papers in Psychology and Economics by Richard Herrnstein*. Cambridge, MA: Harvard University Press.

⁵¹ Rachlin, H. & Green, L. (1972). Commitment, choice, and self-control. *Journal of the Experimental Analysis of Behavior* 17: 15-22.

⁵² Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and self-control. *Psychological Bulletin* 82: 463-496.

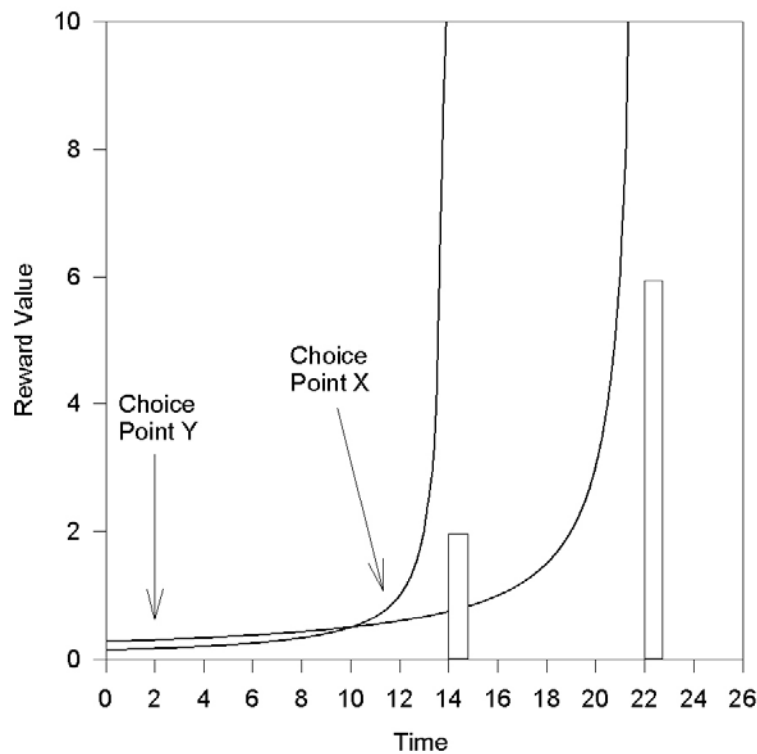
This prediction is that preference between the two rewards will depend on the temporal distance from them at which the choice occurs. At some choice points the individual will be impulsive (choose the SSR) and at other choice points the individual will be self-controlled (choose the LLR), even though nothing differs between the choice points except for the temporal distance from reward. Moreover, this predicted *preference reversal* is in sharp contrast to more traditional views of changes in reward value over time, as we discuss shortly.

Temporal discounting. These choice dynamics are best described with a hypothetical but concrete example⁵³. Say choice of option one (the LLR) leads to a 6-unit reward ($A_1 = 6$) and choice of option two (the SSR) leads to a 2-unit reward ($A_2 = 2$), and that the LLR is available 8 time units after the SSR. (The reward units are measures of value – you can think of them in terms of how much a subject would pay for the reward. The time units can be any regular unit of time – one hour, three days, five seconds, etc.) These relations are shown in Figure 2, with time along the horizontal axis, reward value along the vertical axis, and the rewards represented as vertical boxes at the times they are available. In Figure 2, the 2-unit reward (A_2) is available at time 14, and the 6-unit reward (A_1) is available at time 22. As derived from Equation 1, the value of any delayed reward at any point in time along the horizontal axis is given by dividing the reward amount by the reward delay at that point (A/D), as shown in Equation 2.

$$v_i = \frac{A_i}{D_i} \quad (2)$$

⁵³ Rachlin, H. & Green, L. (1972). Commitment, choice, and self-control. *Journal of the Experimental Analysis of Behavior* 17: 15-22.

Figure 2: Choice Between SSR and LLR



In Equation 2, v_i is the present value of a reward of amount A_i that is available after a delay of D_i . In Figure 2, the curves to the left of the rewards were drawn using Equation 2; they trace the respective A/D fractions for the SSR and LLR as delay changes. Thus, these curves represent the value of the rewards during the times before they are available. Importantly, the reward with the highest value curve *at the time the choice is made* will be preferred. This change in reward value with delay is termed temporal discounting, because the present value of a delayed reward is discounted below what its value will be at the time it is received in the future. Equation 2 is one form of a temporal discounting function, because it describes how reward value changes with delay.

In Equation 1, the B_1/B_2 ratio measures whether the subject prefers option 1 or option 2: a behavior ratio greater than 1.0 indicates a preference for alternative A_1 , while a behavior ratio less than 1.0 indicates a preference for alternative A_2 . At Choice Point X at time 12 in Figure 2, the LLR is delayed by 10 time units ($A_1 = 6$, $D_1 = 10$), and the SSR is delayed by 2 time units ($A_2 = 2$, $D_2 = 2$). Inserting these amount and delay values into Equation 1 yields a B_1/B_2 ratio of .60, indicating a preference for option 2, the SSR (impulsiveness). At Choice Point X, the absolute values (from Equation 2) of option 1 and option 2 are .600 and 1.000, respectively. But preference is predicted to change if we further delay both rewards by 10 time units, so that choice occurs at a greater temporal distance from both rewards. This is shown as Choice Point Y at time 2 in Figure 2. Now option one is 6 units of reward delayed by 20 time units ($A_1 = 6$, $D_1 = 20$), and option two is 2 units of reward delayed by 12 time units ($A_2 = 2$, $D_2 = 12$).

Inserting these modified amount and delay values into Equation 1 yields a B_1/B_2 ratio of 1.80, indicating a preference for option one, the LLR (self-control). At Choice Point Y, the absolute values of alternative one and alternative two (from Equation 2) would be .300 and .167, respectively. Thus, the values of both rewards are less at Choice Point Y than at Choice Point X, because of temporal discounting, and the shape of the discount function inherent in the matching law predicts that preference between the rewards will reverse simply with the passage of time. For example, someone offered a choice between 1 pizza at 1:00 PM 7 days from now and 2 pizzas at 1:00 PM 8 days from now probably would prefer the 2 pizzas delayed by 8 days. However, at 12:55 PM on the 7th day, if the same person were offered a choice between 1 pizza in 5 minutes and two pizzas in 24 hours and 5 minutes, their preference may have reversed and, if so, they would choose the more immediate single pizza.

This research derived from the matching law has revealed that irrationality, in the strict sense of the economist, is in fact normal! This is critically important and worth repeating: Matching law research revealed that both animals and people are naturally disposed to have inconsistent preferences for future rewards as time passes towards the availability of those rewards. At time A they prefer a larger but temporally more distant reward, but at time B when a smaller but sooner reward is available, their preference reverses and they choose the smaller reward. In the hypothetical example in Figure 2, our subject preferred the LLR at choice point Y but later this preference reversed and our subject preferred the SSR at choice point X. This is just like the problem drinker or disordered gambler who says she wants to indulge less so as to have a happy marriage, but then later changes her mind and drinks or gambles. According to this work, having inconsistent preferences over time is the normal state of affairs for both animals and people. Thus, the question of why we sometimes behave irrationally perhaps should be replaced with the question of how we ever can ever resist imminent temptations and behave in ways that are consistent with our long-term interests.

Subsequent research⁵⁴ revealed inadequacies with the simple A/D fraction in Equation 2 as a generally useful temporal discounting function, and led to the slightly modified version in Equation 3.

$$v_i = \frac{A_i}{1 + kD_i} \quad (3)$$

In Equation 3, v_i , A_i , and D_i represent the present value of a delayed reward, the amount of a delayed reward, and the delay of the reward, respectively. The “1” in the denominator of Equation 3 prevents the infinite rise in reward value when delay is zero, as in Figure 2. The k parameter is a constant for any individual that is proportional to the degree of temporal discounting, with higher and lower k values describing greater and lesser degrees of discounting, respectively. Thus, an individual with a higher k value would discount delayed rewards more than an individual with a lower k value; the former individual therefore would be more impulsive than the latter individual. As discussed below, such individual differences in temporal discounting are critical to understanding substance abuse and addiction and disordered gambling.

⁵⁴ Mazur, J. (1987). An adjusting procedure for studying delayed reinforcement. In M. Commons, J. Mazur, J.A. Nevin, and H. Rachlin (Eds.). *Quantitative Analysis of Behavior (Vol. 5): The Effect of Delay and of Intervening Events on Reinforcement Value*. Hillsdale, NJ: Lawrence Erlbaum Associates.

The discovery that hyperbolic temporal discounting was natural meant a significant departure from previous notions regarding how the value of delayed outcomes changed over time. This isn't surprising: previous theory had supposed that people were basically rational (in the economist's sense) and as we've seen hyperbolic discounting is irrational. Dating back to Samuelson's 1937⁵⁵ discounted utility model, both economics and psychology had assumed that discounting would conform to an exponential function, as in Equation 4.

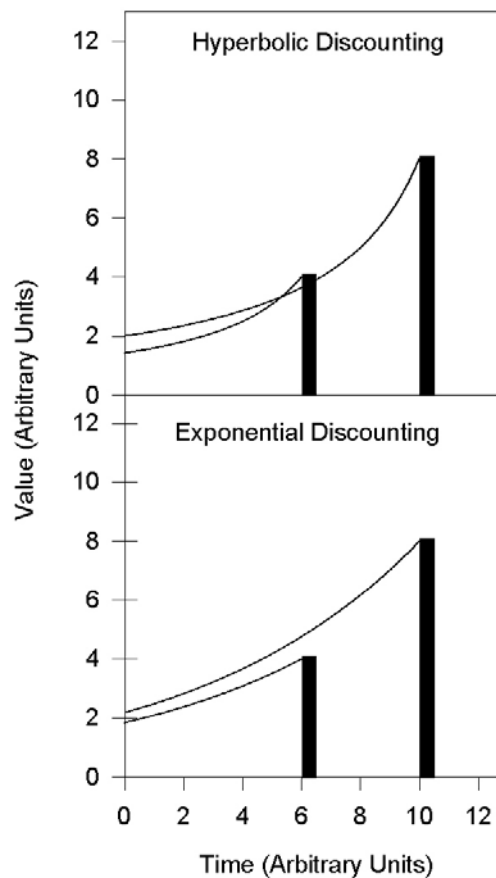
$$v_i = A_i e^{-kD_i} \quad (4)$$

In Equation 4, v_i , A_i , k , and D_i are the same as in Equation 3, and e is the base of the natural logarithms. The key difference between exponential and hyperbolic temporal discounting is that the rate of discounting is constant at all delays in the former but varies with delay in the latter. That is, with an exponential discounting function, equal increments in delay produce constant proportional decrements in reward value. With a hyperbolic discounting function, in contrast, equal increments in delay produce larger decrements in reward value at short delays than at long delays. This critical difference is illustrated in Figure 3. Again, in Figure 3, time is along the horizontal axis, reward value is along the vertical axis, and the rewards are the vertical bars at the times they are available. The top and bottom panels show hyperbolic discounting and exponential discounting, respectively.

The curves to the left of the rewards represent their value during times before they are available. The hyperbolic and exponential curves were drawn using Equations 3 and 4, respectively. The key difference is in the shape of the curves. The exponential curves show a constant change in slope, but the hyperbolic curves are relatively flat at the left of the figure then rapidly rise as the rewards become closer in time. It was quite surprising and initially disconcerting to learn that normal human choice is irrational, as in the hyperbolic curves, and not rational, as in the exponential curves. But now we are armed with that knowledge and so are in a much better position to discover means to get people (including ourselves) to behave in ways that are consistent with their long-term interests and not their short-term interests.

⁵⁵ Samuelson, P. (1937). A note on the measurement of utility. *Review of Economic Studies* 4: 155-161.

Figure 3: Hyperbolic and Exponential Discounting



1.5. Temporal discounting, substance use, and addiction

The evidence is now clear that temporal discounting of reward value is better described by the hyperbolic discounting function than by the exponential function.⁵⁶ These choice dynamics obviously are important for our understanding of addiction. The SSR and LLR are analogous to an alcohol or drug consumption episode and a more valuable but delayed non-drinking or non-drug activity, respectively. The choice dynamics that result from hyperbolic discounting are consistent with the ambivalence that is a critical feature of substance abuse and addiction patterns. That is, even when reward availability does not change, individuals will show ambivalence over time, and their preference for substance use will vary depending on the temporal distance to the availability of the substance and the alternative activity.

⁵⁶ For a review see Kirby, K.N. (1997). Bidding on the future: Evidence against normative discounting of delayed rewards. *Journal of Experimental Psychology: General*, 126, 54-70.

The LLRs will be preferred before the point where the reward value curves cross, and substance use will be preferred after that point. But, once the person exits the situation involving imminent substance use, preference will revert back to the LLRs, and the individual may regret the substance use episode. Such ambivalence is a key feature of addiction.

Significant advances have occurred over the last 10 years in our ability to measure temporal discounting in humans. The basic procedure involves having subjects repeatedly choose between hypothetical money amounts, with a smaller amount available immediately and a larger amount available only after a delay. For example, we start by giving the subject a choice between R10 now (SSR) and R1000 in a week (LLR). Most, but probably not all, subjects would choose R1000 in a week. Then we give them a choice between R50 now and R1000 in a week, and again they probably would choose the LLR.

We continue gradually increasing the amount of the SSR until at some point the subject's preference will change and they choose the SSR. Say the subject chose R1000 in a week over R780 now, but then chose R800 now over R1000 in a week. Given this preference switch, we can reasonably infer that, for this subject, R790 now is equal in value to R1000 in a week. For this subject, R790 is termed the *subjective equivalence point*, or sometimes the 'indifference point', for R1000 in a week, since the subject apparently is indifferent between R790 now and R1000 in a week. Then the procedure is repeated with the R1000 available at several increasing delays, such as one month, six months, one year, five years, etc. A specific subjective equivalence point is thereby found for each of the delays.

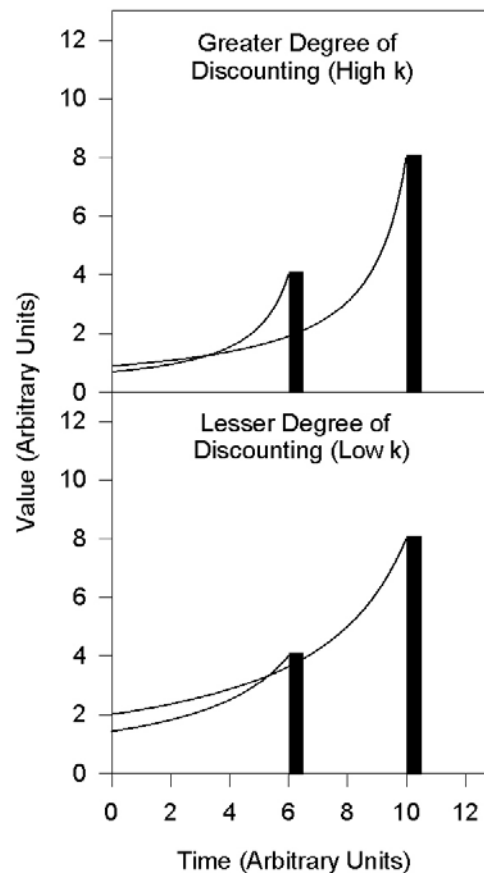
Figure 4 shows the equivalence points for two actual subjects in a discounting study done by one of us.⁵⁷ In Figure 4, the delay of the LLR is along the horizontal axis, and the money amount of the subjective equivalence points is along the vertical axis. These two subjects were chosen because they have very different degrees of temporal discounting, with the filled points showing little decrease in value even at long delays, and the open points showing dramatic decreases in value at even short delays. The equivalence points are subjected to a routine statistical procedure that 'fits' the data to Equation 3, the discounting function, which results in an estimate of the k parameter for each subject. (Fitting a function to data involves finding values in a function so that values produced by the function are close to those in the actual data. In this case, assuming equation 3 to describe the data, fitting is a matter of finding a k value such that equation 3 describes a line passing close to the set of indifference points for a given subject.) The k values for these two subjects are shown in the figure.

BE makes a straightforward prediction regarding differences in temporal discounting which again shows an SSR that is available at time 6 and an LLR that is available at time 10. The reward value curves in both panels of Figure 5 were drawn using the hyperbolic function in Equation 3, with the top and bottom panels showing the curves for relatively high and relatively low values of k , respectively. The higher k used in the top panel produces steeper value curves than the lower k used in the bottom panel. Preference between the SSR and LLR reverses with the passage of time in both panels, with higher and lower discounting producing sooner and later preference shifts, respectively.

⁵⁷ Vuchinich, R.E. & Simpson, C.A. (1998). Hyperbolic temporal discounting in social drinkers and problem drinkers. *Experimental and Clinical Psychopharmacology* 6: 292-305.

As suggested by Figure 5, an individual with a high degree of temporal discounting (top panel) would spend more time preferring the SSR (i.e., substance use) than an individual with a low degree of discounting (bottom panel). This leads to the prediction that k values would be larger among substance abusers compared to normal individuals. This prediction has been supported in several studies of temporal discounting that compared normal individuals with problem drinkers, heroin addicts, smokers, cocaine addicts, and disordered gamblers.⁵⁸

Figure 5: Greater and Lesser Discounting



Summary. BE research on substance abuse and addiction has produced significant advances in our knowledge about the nature of these disorders. We now know that demand for addictive substances can be analyzed with the same economic concepts and methods as other commodities. We now know from the matching law research that inconsistent preferences over time between and SSR and an LLR is the normal state of affairs for humans. And we now know that one aspect of individuals with substance abuse or addiction problems is that they discount the value of future rewards much more so than do individuals without those problems.

⁵⁸ For a review see Bickel, W.K. & Marsch, L. (2001). Toward a behavioral economic understanding of drug dependence: delay discounting processes. *Addiction* 96: 73-86

These basic empirical facts have provided a solid foundation for the development of a number of behavioral theories of addiction, which are summarized in the next section.

1.6. Three BE theories of addiction

The three theories briefly described in this and the next section all take the phenomenon of matching behavior, and especially hyperbolic temporal discounting, as their starting point. They differ in details of how broadly they take the scope of matching to be, what other considerations are important for addiction, and over whether there is anything particular to be said in the case of disordered gambling.

Ainslie's Theory of Addiction Based on Hyperbolic Discounting. Far more than any other author, behavioral economist and psychiatrist George Ainslie⁵⁹ has vigorously and extensively pursued the implications of the preference reversals caused by hyperbolic temporal discounting for understanding impulsiveness and impulse control in addiction, as well as in obsessive-compulsive disorders, dissociative disorders, bulimia, itches, physical pain, and flaws of character.

Hyperbolic discounting can explain impulsive choices that are later regretted, and thus can explain the addict's repeated surrender to temptation, despite earlier resolutions to quit or curtail substance use. That is excellent progress as far as it goes, but so far it does not explain how such impulsiveness can ever be avoided. How do addicts and the rest of us at various times resist temptation and exercise self-control? More precisely, why is the LLR ever preferred to the SSR, especially when the SSR is immediately available? Ainslie has described four processes that can be used to subvert preference reversals and thus can lead to choices that are more consistent with our long-term interests.

External Commitment. The most obvious self-control tactic is *precommitment*, meaning the deliberate engineering of the physical or social environment so that the impulsive choice becomes impossible or much more difficult to make. In Figure 2, for example, the LLR is preferred at time 2 but the SSR is preferred at time 12. An individual with this preference pattern could exercise self-control by doing something at time 2, when the LLR is preferred, that makes it impossible or very difficult to reverse course and choose the SSR when it becomes preferred at time 12. This device is exemplified by the myth of Ulysses and the Sirens, in which Ulysses chained himself to the mast of his boat before sailing past the deadly temptation of the sirens, whose charms he knew he'd be unable to resist if he left himself free to succumb. Notice that Ulysses thereby showed awareness of his own disposition for irrational preference reversal, and used this foreknowledge to block its dangers.

⁵⁹ Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and self-control. *Psychological Bulletin* 82: 463-496; Ainslie, G. (1992). *Picoeconomics: The strategic Interaction of Successive Motivational States Within the Person*. Cambridge, England: Cambridge University Press; Ainslie, G. (1999). The dangers of willpower. In J. Elster and O.-J. Skog (Eds.), *Getting Hooked: Rationality and Addiction*. Cambridge, Cambridge University Press: 65-92.

An obvious example from the addictions field is the voluntary ingestion of the drug disulfiram, which a dependent problem drinker knows will make her feel very ill if she consumes alcohol. Interestingly, even pigeons will make precommitments of this sort under carefully set up conditions⁶⁰.

Control of attention. This refers to restricting the processing of information about the availability of the SSR. It is somewhat similar to the Freudian concept of repression, although originating obviously from a very different theoretical base. From Ainslie's viewpoint, the avoidance of information, unlike in repression, can be conscious as well as unconscious. For example, a recovering alcoholic may avoid walking past her usual pub. In Chapter 3, we will see reasons for thinking that this is probably an essential aspect in the control of any full-blown addiction once it has developed.

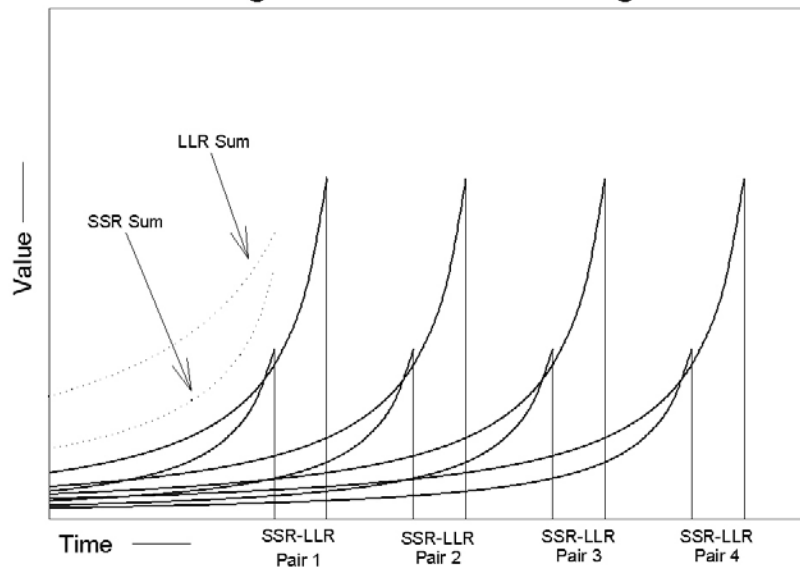
Preparation of emotion. This describes attempts to inhibit emotional responses associated with the SSR or to augment emotions that are incompatible with the appetite for it. In the case of addictions, it mainly applies to the cognitive control of so-called 'craving' responses. For example, people trying to quit smoking may absorb themselves in pictures of diseased lungs, or put the money they had been spending on cigarettes in a kitty used to reward themselves with periodic indulgences, so as to make vivid to themselves the financial costs of smoking in addition to its health costs.

Personal rules. This process is by far the most important category of methods of self-control described by Ainslie, and the one that is used to create a major extension to theory. This part of Ainslie's work provides a BE account of *willpower*, that mysterious force that is widely believed to be responsible for how and when we resist temptation and behave according to our long-term interests. We will explain personal rules at greater length than the other devices, both because this idea will feature most prominently in our policy discussion in Chapter 4, and because it involves more complexities.

The basic case of temporary reversals of preference due to hyperbolic discounting involves a single SSR-LLR pair, as in several of the previous figures. But, of course, in life outside the laboratory many identical or very similar SSR-LLR choices will occur repeatedly in the future. Ainslie's personal rules concern whether and when an individual will perceive each SSR-LLR pair as a separate choice, or whether she will perceive the choice as being between a whole series of SSRs and a whole series of LLRs. If the individual perceives the choice as being between whole series of SSRs and LLRs, then she is engaging in *reward bundling*, a potentially powerful self-control process. Reward bundling is illustrated in Figure 6. In the figure, time is along the x-axis and reward value is along the y-axis. The figure shows four SSR-LLR reward pairs spread out at the times they will be available. The solid curves to the left of the rewards were drawn from Equation 3 and represent the value of the rewards during the times before they are available.

⁶⁰ Rachlin, H. & Green, L. (1972). Commitment, choice, and self-control. *Journal of the Experimental Analysis of Behavior* 17: 15-22.

Figure 6: Reward Bundling



If the person facing this series of choices takes the SSR-LLR pairs one at a time, then we would see repeated preference reversals. At the far left of the figure, she prefers the Pair 1 LLR, but when the Pair 1 SSR is imminent, its value rises above that of the Pair 1 LLR, so her preference reverses and she chooses the SSR. After Pair 1 is past she faces Pair 2. Initially the Pair 2 LLR is preferred, but again as the Pair 2 SSR becomes imminent its value rises above that of the LLR and her preference once again reverses. The same thing then would happen for SSR-LLR Pairs 3 and 4. This is a common pattern for individuals with impulse control problems. Many people wake up each morning preferring not to do what they did yesterday, not to drink or smoke or gamble or indulge in whatever other self-control problem they have, but then later in the day as the object becomes available they succumb to temptation.

On the other hand, if the person facing this series of choices did not take them one at a time, but instead viewed their situation as one single choice between the whole series of SSRs and the whole series of LLRs, then we could see much more self-control. In Figure 6, the reward value curves for all four SSRs and all four LLRs are summed during the times before the Pair 1 SSR is available. These summed value curves are the dotted lines at the left. The key point is that the LLR sum is greater than both the SSR sum and the value of the Pair 1 SSR at the time the Pair 1 SSR is available. So, an individual who viewed this as one choice between series of rewards would prefer the LLR series even when the SSRs are immediately available. This also is a common pattern for individuals *without* (major) self-control problems. Many people wake up each morning not even thinking about drinking, smoking, or gambling, because they didn't do it yesterday, or the day before yesterday, or the day before that; some time ago they made the choice not to engage in the SSRs and so now are reaping the greater benefits of the LLRs.

To avoid a reversal of preference, a person can *bundle* a whole series of choices together in anticipation of the higher aggregate reward that would be obtained from preferring the LLRs. She does this by adopting *personal rules* that dictate the choice to be made in a whole class of conflict situations involving the need to delay gratification.

These personal rules correspond to the idea of ‘principles’ or ‘universal rules’ of behavior that have been recommended as a means of achieving will-power by philosophers down the ages. The person is seen as committing themselves to a course of action by making *private side-bets*, a form of personal rule-making in which the current choice serves as the best predictor of what future choices will be. If the current SSR is successfully resisted, the bet is won, and the expectation of future reward is proportionately strengthened and the ability to overcome similar temptations in the future is enhanced. On the other hand, if the SSR is not resisted, the bet is lost, and also lost is whatever elements of her view of herself had been staked, with a decreased expectation of being able to resist the SSRs of the same kind in future.

Personal rules and private side bets can be illustrated with Figure 6. Say the individual has a strict rule of “only go to the bar for a few drinks on Fridays and legal holidays.” If followed, this rule would allow enjoyment of drinking on the designated days but would also produce choices of the LLR over the SSR (drinking) the rest of the time. But when the Pair 1 SSR is immediately available on a regular weekday, there is nothing in the physical or social environment that prevents her from choosing the SSR or that even makes choosing the SSR much more difficult. What is at stake in the private side-bet is the difference between the summed LLR and SSR value curves at the time the Pair 1 SSR is available, because the reward that is chosen from Pair 1 strongly predicts the reward that will be chosen from subsequent pairs. If the first SSR is resisted, she gets the first LLR plus the enhanced expectation of also getting subsequent LLRs reflected in the LLR sum. On the other hand, if the first SSR is not resisted, she gets that SSR but also gets a reduced expectation of getting any of the subsequent LLRs. Because of the predictive value of current choices for later choices, *which is a reward the chooser gets right now as she makes the choice*, personal rules can bring into the present the value of rewards in the future, and thus allow people to overcome the effects of their natural hyperbolic discounting.

Take, for example, an overweight person who wants to lose some weight primarily for health reasons. If he thinks truly seriously about his situation, he can notice that when he chooses to fix or not to fix himself another sandwich, he isn’t just choosing between satiation now and sticking with his diet today. If he prefers to be trim in the distant future, his choosing a sandwich now provides him with evidence that he won’t choose his diet then either; for hyperbolic discounters, present choices predict future choices. In light of this, perhaps he can frame his present choice to reflect the fact that it isn’t just between immediate streams of benefits and costs, but has implications all the way along the flatter part of his discount curve into the future. If all those preferred states of fitness in the future are at stake now, then their value, added together, might be enough to trump the present value of the sandwich. It won’t typically be enough for the person just to tell himself this; if it were, self-control would be easy and addiction and impulsiveness would both be mysterious. The would-be dieter needs to arrange his patterns of action so as to make it impossible for him to avoid keeping the future implications of his action in mind while he makes his present choice. Suppose he therefore frames a personal rule to the effect that an extra sandwich is allowed only on a day after his visits the gym. Then, if today isn’t a day after a gym session, his choice is between a sandwich now and now facing the long-term ruin of his personal rule. The key thing to see here is that because present choices predict future choices the personal rule is a presently valuable asset. It bundles a whole series of future benefits and puts them at stake in the present moment. If he throws his rule away by breaking it, he will suffer from his knowledge of the loss of it here, where discounting is at a minimum.

Ainslie goes on to propose that, rather than being characterized by rational and consistent preferences as conventionally assumed, the self is in fact made up of conflicting preferences for a variety of rewards that, because of hyperbolic discounting, achieve dominance over each other at different points in time. Each appetite for and motivation to acquire each reward is described as an 'interest' in that the reward and the interaction between these competing interests is portrayed as an internal economic market-place. In particular, according to this model, the interests of a present self compete strategically with those of an anticipated future self that may attempt to subvert them. This battle may be complicated by the existence of mid-range interests against which short- and long-range interests may combine forces. This strategic interaction of interests can be illustrated by further complicating the example earlier of the overweight person. In this example, the short- and long-range interests are the extra sandwich now and being slim and healthy in the future, respectively. But there are other interests at intermediate ranges between short and long that may interact with the short- and long-range interests and impact the critical choice to have or not to have that extra sandwich. Say, for example, he has a mid-range interest in following his favorite football team, and another mid-range interest in getting dates with attractive women. Say on a day he didn't go to the gym, his team happens to be playing a televised game. In this situation, his extra sandwich interest may recruit his sports interest and both together then overcome the future-slim-self interest, and he chooses to watch the game and eat the extra sandwich. Alternatively, say on a day he didn't go to the gym, he meets an attractive woman but makes no overtures because he is overweight. Later when he gets home he will be tempted to make the extra sandwich, but the future-slim-and-healthy-self interest could recruit the attractive date interest, and both together overcome the extra sandwich interest, so he resists the extra-sandwich temptation and gets closer to his intermediate and long range goals of dating attractive women and being slim and healthy.

So how are relatively stable choices ever made in this intrapsychic free-for-all of successive motivational states attached to short-range, mid-range, and long-range interests? Ainslie here borrows concepts used to analyze interpersonal or international conflict⁶¹ and applies them to the intrapersonal bargaining situation.⁶² In particular, he uses the model of repeated instances of the 'prisoner's dilemma' game, in which players must choose between cooperative and non-cooperative options, to represent a bargaining process involving successive motivational states.

Here is a version of the standard illustration of the prisoner's dilemma: Two men commit a burglary and are arrested while in possession of burglar's tools. Possession of the tools is a minor crime. The police have no evidence against the men, other than the tools, so they offer each man the following incentives. If neither man confesses, they won't be convicted for the burglary, but both will get a 90-day sentence for the tools. If both confess to the burglary, they'll both be convicted and both will get 2-year sentences. If one confesses and the other doesn't, the confessor will go free and the non-confessor will get a 5-year sentence. The dilemma is: should each man do what is best for himself alone, or should he do what is best for both men as a pair? For each thinking only of himself, the best thing to do is confess, since he'll then get his best possible outcome - going free if the other doesn't confess or a 2-year sentence (which beats 5) if the other does confess – regardless of what the other does.

⁶¹ Schelling, T.C. (1960). *The Strategy of Conflict*. Cambridge MA, Harvard University.

⁶² Ainslie, G. (2001). *Breakdown of Will* Cambridge University Press: Cambridge, UK, p. 90.

The other prisoner reasons in exactly the same way, since his situation is identical. So they both confess and get 2-year sentences. But now notice that when we think of the pair as a whole, they're both worse off than they would have been had neither confessed – then they'd have each got just 90 days. Ainslie has extended such *interpersonal* (i.e., between people) strategizing to consider the *intrapersonal* (i.e., within the person, between interests) bargaining we all engage in over time. Rather than two people at the same time considering what is best for themselves alone or what is best for the pair, we now have one person at different times considering what is best for the present self alone or what is best for both the present self and the future self as a pair. When the present self confronts an SSR-LLR pair of rewards, the best thing for the present self alone is to choose the SSR, because only the SSR is available at the present time. On the other hand, the best thing for the present self *and* the future self together as a pair is to choose the LLR, because then over time they'll receive the greater benefits of the series of LLRs.⁶³ This intricate and subtle arena of competing personal rules and complex intrapersonal bargaining strategies leaves plenty of room for the evasions, distortions and other forms of self-deception well-recognized in the addict's behavior, as well as in ordinary human conduct, and these are deduced by Ainslie from his theory. In Chapter 4 we will discuss the implications of these ideas, especially ways to promote reward bundling, for dealing with gamblers.

Two molar BE theories of addiction. The temporal inconsistencies in preference caused by hyperbolic temporal discounting are a critical element in explaining addiction from the BE perspective. Hyperbolic discounting alone, however, is widely regarded⁶⁴ as an insufficient explanation. Thus, choice dynamics other than hyperbolic discounting have been proposed to explain the apparently self-defeating behavior pattern of addiction. In this section, we sketch two molar theories (see Section 1.2 above) of addiction that have been developed within the BE perspective: (1) Richard Herrnstein's melioration theory of addiction⁶⁵, and (2) Howard Rachlin's relative theory of addiction⁶⁶.

Both these theories accept hyperbolic discounting as fact, so both are entirely consistent with Ainslie's work. However, these theories focus on an aspect of consumption other than hyperbolic discounting.

⁶³ In the prisoner's dilemma between people, it is simply a mathematical fact, having nothing to do with psychology, that the only achievable outcome for rational players is for each prisoner to do is confess, despite the fact that this makes the best joint outcome unachievable. The basic cause of this is that, in the situation as stipulated, neither prisoner can do anything to give the other a reason to stick to a bargain. So how do selves prevent their inner prisoner's dilemmas from breaking down in this way? In the prisoner's dilemma, mutual confession is no longer rational *if* the prisoners expect to find themselves in the same situation again, since then each can promise to keep quiet, and each can threaten to punish the other one next time if he breaks his word. As long as the punishments are severe enough, now each can rationally refuse to confess and they can both get their light sentences. Similarly, Ainslie argues, the interests within a self are 'locked up in a room together' and have to interact in the future. Thus they too can sometimes cooperate.

⁶⁴ Monterosso, J. & Ainslie, G. (1999). Beyond discounting: Possible experimental models of impulse control. *Psychopharmacology* 146: 339-347.

⁶⁵ Herrnstein, R.J. & Prelec, D. (1992). A theory of addiction. In G. Loewenstein and J. Elster (Eds.), *Choice Over Time* (pp. 331-360). New York: Russell Sage Foundation.

⁶⁶ Rachlin, H. (1997). Four teleological theories of addiction. *Psychonomic Bulletin and Review* 4: 462-473.

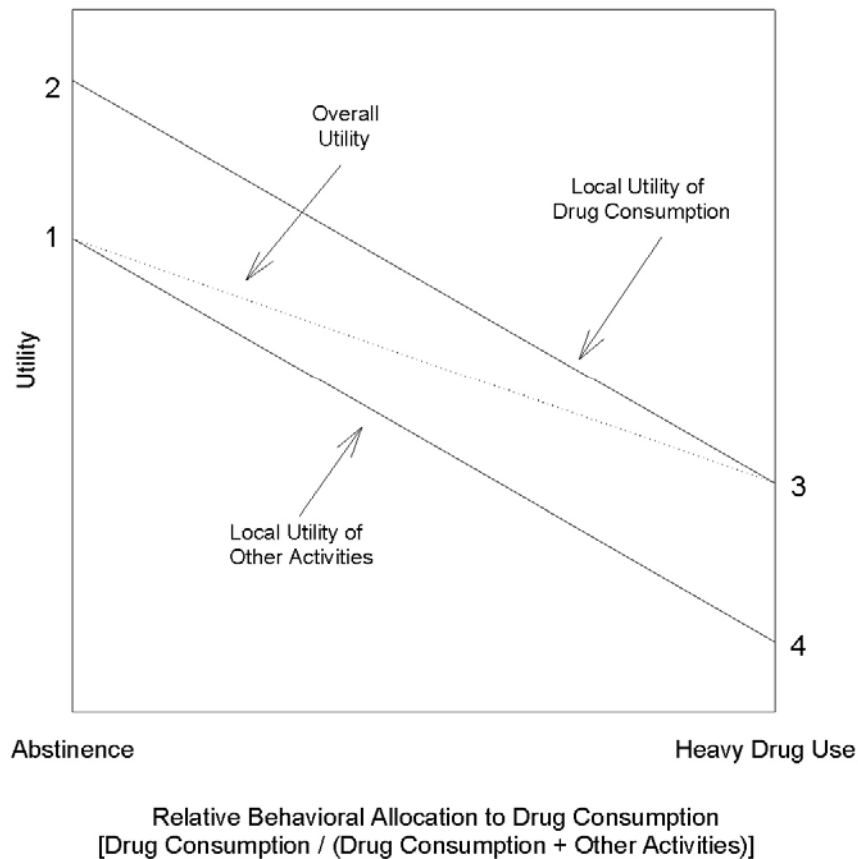
That aspect is that consumption of addictive commodities during one time period reduces the utility derived in later time periods both from the addictive substance and from engaging in other activities. In general, utility is defined as the benefit derived from engaging in an activity. Because of tolerance, consumption of the addictive commodity in one time period reduces utility of the same level of consumption in future periods. For example, if an individual consumed two drinks every night over several weeks, she would get a pleasant high from the two drinks early in the period but the two-drink high would gradually diminish until by the end of the period the two drinks produced little effect. For this reason, regular drinkers must either be satisfied with diminishing benefits from drinking, or, as more often happens, increase their drinking over time to obtain the same high. Moreover, because consuming addictive substances can disrupt one's ability to engage in and enjoy alternative activities, consumption of the former in one time period reduces the utility derived from the latter in later time periods. Thus, these theories propose a sort of intertemporal dependency between consumption and utility in different time periods that is critical in accounting for addiction. One final preliminary note: a key aspect of both theories is the difference between local utility and global utility, a distinction made earlier in this chapter (Section 1.2). Local utility refers to the utility one gets immediately from an action, while global utility refers to the overall utility one gets from a given sequence of actions.

Herrnstein's melioration theory of addiction. Because of the process of melioration (see Section 1.2), in addition to hyperbolic discounting, an individual's choices are relatively insensitive to global utility and much more sensitive to the comparison of local utilities. Herrnstein's theory shows how this process can lead to addiction. As noted above, a key element of the melioration theory is that substance use lowers the utility derived from such use and also from non-substance use activities. This process is depicted graphically in Figure 7, in which utility is along the vertical axis, and the relative choices of substance use and of non-substance use activities is along the horizontal axis. The farther to the right along the horizontal axis, the more the individual is choosing substance use and not choosing other activities – the process of addiction – and conversely for movement to the left along the horizontal axis. So, at the far right, the individual is consuming the substance and doing little else, and at the far left, she is doing a lot of other things and not consuming the substance. The local utility of substance use, the local utility of other activities, and global utility are signified by lines 2-3, 1-4, and 1-3, respectively, at the various distributions of choices along the x-axis. The melioration process dictates that choices are made according to which choice option provides the highest local utility at the time. As seen in Figure 7, even though both local utilities and global utility are being reduced by substance use, the individual continues to use the substance. Increasing levels of addictive commodity consumption, movement to the right along the horizontal axis of Figure 7, reduce both local and global utility, but the melioration process continues to drive addictive consumption because the local utility of that consumption remains higher than the local utility of not consuming.

Theoretically, the process of recovery from addiction may involve a restructuring of how the choice alternatives are perceived. The above discussion assumed that only events during the substance use episode are included in the computation of local utility derived from substance use. So, the utility derived from an evening of heavy drinking is kept separate from the next morning's hangover. If, however, the temporal boundary of what counts as local utility is expanded, this may alter the computation of local utility so that the value of a substance use episode is changed. In this example, if local utility is expanded to include the time spent drinking plus the hangover, and not just the drinking, this may reduce the utility derived from drinking and therefore lower its value.

Notice that this is bundling again. From one BE model to another, successful bundling keeps emerging as the crucial defense against addiction.

Figure 7: Melioration Addiction



Rachlin's relative theory of addiction. Rachlin's model is similar to and entirely consistent with Herrnstein's. The key difference is that, whereas Herrnstein grouped all activities other than substance consumption into one large category, Rachlin focuses on a particular class of activities involving *social interaction*. Rachlin's theory is based on two important aspects of social interaction:

First, social interaction is a close economic substitute for various addictive substances (i.e., smoking, drinking, drug use). When two commodities are substitutes, their consumption varies inversely; the more of one that is consumed, the less of the other that is consumed (and vice versa).

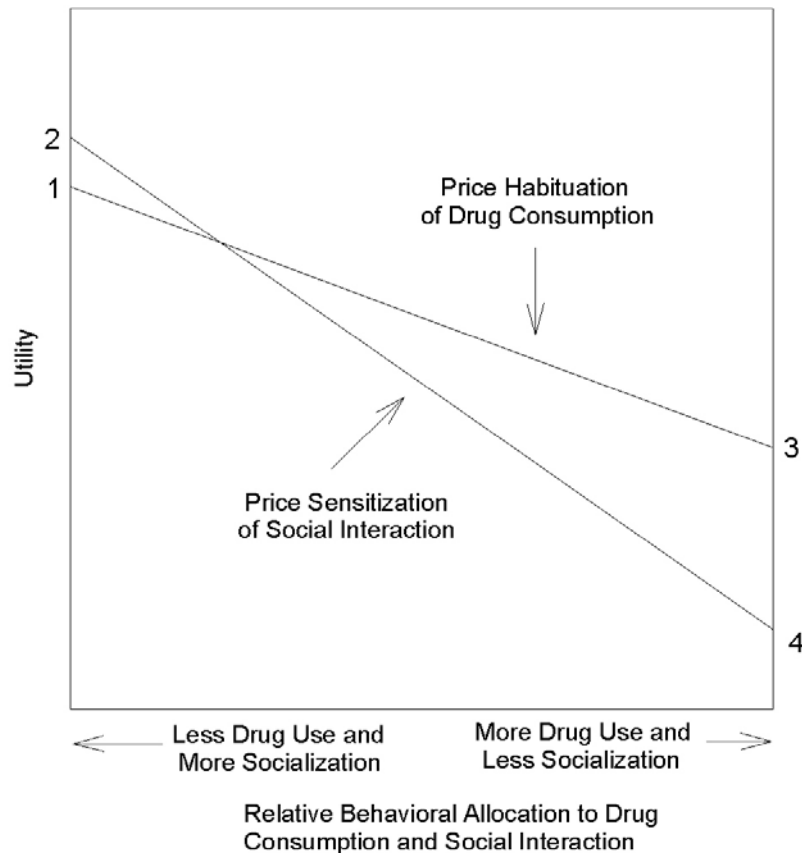
Second, the utility derived from substance consumption and from social interaction changes in opposite ways as these activities are engaged in. We noted earlier that, because of tolerance, the utility derived from a given level of substance consumption decreases over time. On the other hand, the utility derived from a given amount of social interaction increases over time. So, if you have two drinks tonight, those two drinks would provide much more utility if you haven't had any alcohol for several weeks than if you've been having two drinks every night for several weeks. Conversely, if you encounter a stranger at a bus stop and converse for five minutes, that brief interaction would provide more utility if you'd recently been socially interacting regularly than if you'd recently been socially isolated.

Rachlin termed those changes in the utility derived from substances and from social interaction the *price habituation* of substance consumption and the *price sensitization* of social interaction. He framed this in terms of price changes because a given level of utility costs more or less over time. The more a substance is consumed, the higher the price for a given unit of utility (and vice versa). Conversely, the more one interacts socially, the lower the price for a given unit of utility (and vice versa).

How the price habituation of addictive commodities and the price sensitization of social interaction can lead to addiction is illustrated in Figure 8. Like Figure 7, in Figure 8 movement to the left along the horizontal axis signifies less drug use and more social interaction, and movement to the right along the horizontal axis signifies more drug use and less social interaction. The utility derived from a given choice of substance use and social interaction is signified by lines 1-3 and 2-4, respectively, at the various distributions of choices along the horizontal axis. The price habituation and price sensitization of substance use and social interaction, respectively, are shown with lines sloping down to the right. Movement to the right raises the price of both substance consumption and social interaction (a given unit of utility of either costs more because the substance use has been increasing and social interaction has been decreasing. Conversely, movement to the left lowers the price of both substance consumption and social interaction (a given unit of utility of either costs less because the substance use has been decreasing and social interaction has been increasing.) Given the melioration process, the person will choose the option that provides the highest utility at the time the choice is made.

As an example, assume that over a period of time an individual's choices are stable with a pattern of drinking and social interaction at the point of intersection of lines 1-3 and 2-4 in Figure 8. This choice pattern involves relatively more social interaction and less substance use. However, say an event (e.g., marital separation) then occurs that makes social interaction more expensive in the sense that it now is more difficult to obtain, which would reduce the number of choices of social interaction. Because drinking and social interaction are mutually substitutable, after the event the individual begins to drink more to make up for less social interaction (movement to the right along the horizontal axis in Figure 8). Because of the price habituation and price sensitization of drinking and social interaction, respectively, drinking more and socializing less will decrease the utility derived from both drinking and socializing. But even though both utilities are decreasing as the individual moves to the right, the choice dynamic of relative addiction theory continues to drive addictive consumption because the utility from drinking exceeds that of the utility from social interaction.

Figure 8: Relative Addiction



Given its focus on the utilities and costs of substance consumption and social interaction, relative addiction theory would look to these variables as being important in the recovery from addiction. Events that produced a large enough increase in the price of addictive substance consumption and/or a large enough decrease in the price of social interaction could tip the balance of the cost/benefit ratios so that an addicted individual would curtail or quit their consumption.

1.7. Extensions of BE research to gambling

We hope we have now persuaded the reader that BE theory and research has significantly advanced our knowledge of substance abuse and addiction. Because the behavioral manifestations of substance abuse and addiction and of disordered gambling are so similar, if not the same, there is every reason to be confident that BE could make significant contributions to our understanding of disordered gambling as well. (Note that we are careful here to say that the behavioral structure of disordered gambling and addiction are similar.

We will later argue that only a subset of those with pathological gambling have a genuine addiction. But BE's basic model applies to many behavioral patterns – from overeating to frittering away too much time playing internet solitaire – that are also not addictions.)

As applied to disordered gambling, the larger, later rewards (LLRs) would appear to be the same as those contrasted with substance abuse and addiction: adaptive functioning in key life-health areas such as intimate, family, and social relations, and vocational success. The smaller, sooner rewards (SSRs) are different, however, because the gambler takes nothing into her body as does the substance abuser or the addict. Neuroscience is now telling us about this difference in detail, and we will describe these advances in Chapter 3. In any case, the apparent irrationality of the disordered gambler and of the substance abuser or addict are the same: At time A they truthfully report that they want to gamble less or not at all so as to improve their lives in more important ways, but at some future time B that stated preference reverses and they gamble. Then later at time C, when regret sets in, they are confronted again with their inconsistent preferences for a better life and excess gambling.

In this section we will discuss Ainslie's and Rachlin's specific perspectives as applied to disordered gambling. Both models are entirely consistent with each other, but they focus on different aspects of impulsiveness and self-control. Because Rachlin's model of addiction in general is consistent with Ainslie's, and then adds to it a particular account of disordered gambling, all policy and intervention implications of Ainslie's model will carry over to Rachlin's and then be supplemented by it. We should make clear that all that is being attempted at this point is logical extrapolation from theoretical perspectives. We will turn to the implications of direct empirical studies in the succeeding chapters of this report. Also note that in this section we will only briefly mention practical implications, as they will be spelled out in considerably more detail in Chapter 4.

We'll first discuss how Ainslie's model can be used to conceptualize the choices faced by the disordered gambler. As noted earlier, in this case the SSRs are opportunities to gamble and the LLRs are quality of life concerns in other areas. As in Figure 6, the gambler faces a long series of such choices. Any of Ainslie's self-control tactics discussed earlier (external commitment, control of attention, preparation of emotion, and personal rules) could be brought to bear to make current choices more consistent with long-term interests. We will focus, however, on personal rules (reward bundling), because this is by far the most important self-control tactic and the one with the most extensive implications.

An individual could develop personal rules that are in force before she even enters the casino. As with the earlier discussion of rules regarding drinking, a gambler could, for example, generate a rule that states she can only go to the casino on Fridays and holidays. If followed, this rule would allow her to enjoy gambling (the SSR) on a relatively small number of designated days, and also enjoy the greater benefits of the LLRs on days she is not gambling. The more times the rule is followed, the stronger it will become as she develops heightened expectations of continuing to be governed by it in the future and in reaping the benefits of the LLRs. As with all personal rules, however, it will always be vulnerable to being broken if a coalition of interests gets strong enough to subvert it.. If given an opportunity, the short-term interest in gambling could recruit a mid-range interest to its side of the conflict and subvert the rule. Suppose, for example, that on a Friday in the casino the gambler meets an attractive man who asks her if she will be back on Saturday night.

Rigid adherence to the rule would require her to say “no,” which she may well do. But as Saturday night approaches, her mid-range interest in dating attractive men could combine with the short-term interest in gambling again on Saturday, which may be enough to overwhelm the long-term interest, and the rule. If she gambles on Saturday, she would have broken the rule, and in so doing has now reduced her present expectation of rule following in the future and of receiving future LLRs.

Gamblers can also develop personal rules that are in force while they are in the casino. As an example, suppose there are two gamblers, A and B, and that A has no personal rules regarding the extent of his gambling while in the casino. For A, with no rule, there is nothing at the moment that can bring possible future LLRs to bear on his current decisions. With nothing to connect his present self, which strongly prefers the SSRs of gambling, to his future self, which will prefer the LLRs, the SSRs will have their way with his decisions and he will gamble until he runs out of money or the casino closes. On the other hand, suppose B has a rule that states: “I’ll count my money every hour. If I’m losing money at the count I’ll leave the casino. If I’m winning money at the count I’ll stay another hour, but under no conditions will I stay longer than three hours.” At each count B faces the choice of doing what the rule dictates, or continuing to gamble. If he follows the rule, then it becomes stronger and he has increased his present expectation of continuing to reap LLRs in the future; if he breaks it he will bring on the reverse present expectation. If the rule collapses and he fails to generate another, effective, one then he will be in the same situation as gambler A. Of course, like rules in force outside the casino, rules for choices inside the casino can be subverted by a coalition of interests. Suppose at the hour 2 count B is losing money, and the rule enjoins him to leave the casino. But, it happens that 15 minutes ago a valued business contact arrived at the table and had been looking at him expectantly. His short-term interest in continuing to gamble could coalesce with the mid-range interest in networking and together they may subvert the rule.

These examples of applications of bundling in gambling situations have been simple and fanciful. We will discuss more serious measures that could be adopted as public policies to make it easier for people to bundle while gambling, or that could limit casinos’ capacities to inhibit their bundling, in Chapter 4.

Rachlin’s relative addiction theory supplements Ainslie’s emphasis on intrapersonal bargaining with additional, specific implications for gambling behavior. Rachlin assumes that gambling behaves like the consumption of addictive substances in two ways. First, the more one gambles, the less utility one derives from a given amount of gambling. In other words, gambling is price habituated like addictive substances, so there is an inverse relation between the extent of gambling and the utility derived from gambling. Second, gambling is mutually substitutable with social interaction, so the less one gambles the more one interacts socially (and vice versa). If Rachlin is on the right track, then measures should be sought that increase the price of entry to gambling situations for disordered gamblers, and/or decrease their price of entry to situations of social interaction. We will discuss these possibilities more fully in Chapter 4.

As mentioned earlier, Rachlin is the only BE theorist to date who has an explicit and specific model of problem gambling. According to this theory⁶⁷, a person at risk for problem gambling is prey to an insidious perceptual process or accounting scheme.

⁶⁷ Rachlin, H. (1990). Why do people gamble and keep gambling despite heavy losses? *Psychological Science* 1: 294-297; Rachlin, H., Siegel, E. & Cross, D. (1994). Lotteries and the time horizon. *Psychological Science* 5: 390-393.

We will introduce the idea by referring the reader back to our earlier discussion of Herrnstein's point that, the size of the behavioral unit over which utility is calculated is not carved in stone. With drinking, for example, is the utility of drinking calculated by adding up each individual sip, each drink, or each drinking episode? In that same discussion, we also noted that recovery from addiction may involve a process of restructuring the alternatives so that the behavioral unit over which utility is calculated is made larger. An individual whose drinking behavioral unit also included the next day's hangover would be more likely to reduce their drinking than an individual whose behavioral unit included the night of drinking alone.

In the case of gambling, the question of the behavioral unit over which utility is calculated concerns how bets are aggregated into larger units. At one extreme end of a continuum, each bet could be seen as a single, independent event. Alternatively, at the other end of the continuum, each visit to the casino could be seen as a single, independent event (where an 'event' just names the unit used for cost-benefit analysis by the person). Rachlin argues that the disordered gambler sorts her gambles into strings in which each string terminates with a win. That is, she thinks of each win as redeeming the sequence of losses that preceded it, *regardless of how long that sequence was in any given case*. Her losses redeemed, the DG is ready to start a new sequence, with a clean record, following every nod from lady luck. This perceptual or bookkeeping scheme interacts with hyperbolic discounting in an insidious way. Each string the gambler perceives will have the form of a win preceded by n losses where $n \geq 0$. It is the sequences in which n is largest that account for most of the financial distress – this is what makes problem gambling a *problem*. But now: the overall outcomes of long sequences will have their values hyperbolically discounted relative to the elements of short sequences. Long sequences, as noted, are where the losses are disproportionately concentrated. Thus, the gambler weights wins after short strings at closer to their full value than wins after long strings, with the average discrepancy being greater the worse the ratio of wins to losses!

Let us show an example. Ls and Ws below represent sequences of losses and wins, respectively, in a simple game of tossing a coin:

WLWLLWWLLLWLLLWWWWLW

There are 11 wins and 11 losses in the entire string, so as a whole it is neither positive nor negative in value. In the same W-L sequence somewhat modified below, the outcomes are organized as in Rachlin's theory, with each sub-string terminating with a win. The numbers below each win represent the overall outcome for each substring:

W	LW	LL W	W	LLL W	LLL W	W	W	W	LL W	W
+1	0	-1	+1	-2	-2	+1	+1	+1	-1	+1

Summing the outcomes of the substrings again results in zero, so the overall outcome is neither positive nor negative. Importantly, however, because the W that terminates each sub-string is some distance into the future when the string begins, the present value of the overall outcome of the sub-string will be discounted.

Using the hyperbolic discount function in Equation 3, with k equal to 1 and the delay equal to the number of coin tosses in the substring, the actual overall value and the discounted overall value of each sub-string would be:

	W	LW	LL W	W	LLL W	LLL W	W	W	W	LL W	W
Undiscounted	+1	0	-1	+1	-2	-2	+1	+1	+1	-1	+1
discounted	+1	0	-.33	+1	-.67	-.67	+1	+1	+1	-.33	+1

Summing the discounted values of each sub-string yields a value of 4, a decidedly positive overall outcome. Because of this perceptual organization of behavioral units into strings terminating in a win, the disordered gambler perceives positive value when in reality, in this example, the overall outcome is neutral. Because of the perceptual organization and hyperbolic discounting, the disordered gambler fools himself into perceiving positive value where in fact there is none.

This looks like a disastrous way to think, so it is initially surprising that, as Rachlin documents, animal behavior suggests it to be common in nature. Why would natural selection have built so perverse a perceptual quirk? The answer presumably is that in normal environments rewards are distributed in a patchy way. Thus when an animal finds *any* reward in an area, it should persist through some disappointments in searching the region further. But modern gambling environments, unlike any in nature, are built to exploit the bias for following up reward. Thus a disposition that was helpful to our ancestors is harmful to many of us. Nevertheless, this notion suggests that we consider possible policy interventions that encourage disordered gamblers to employ alternative accounting schemes.

1.8. How many kinds of disordered gambling are there?

Disordered gambling (DG) is simultaneously a social problem, a clinical challenge facing individual people and their counselors and physicians, and a scientific puzzle. It is not easy to keep inquiry motivated by all of these considerations from being, on the one side, monopolized by one of these considerations to the detriment of the others, or, on the other side, falling into incoherence. As Howard Schaffer⁶⁸ notes, DSM-IV equivocates on whether it regards pathological gambling (PG) as a 'real' phenomenon when it cautions that "[I]nclusion here, for clinical and research purposes, of a diagnostic category such as Pathological Gambling ... does not imply that the condition meets legal or other non-medical criteria for what constitutes mental disease, mental disorder, or mental disability." Is the motive for this equivocation mainly derived from social policy considerations or scientific ones? There is probably no straightforward answer to this question. But to the extent that there was scientific consensus, it would be bizarre for DSM to continue saying what it does. If there were agreement that pathological gamblers are merely people who are irresolute in the face of bad habits, then PG should be removed from the DSM.

⁶⁸ Shaffer, H. (1999) Strange bedfellows: a critical view of pathological gambling and addiction. *Addiction* 94: 1445-1448.

If, on the other hand, there were consensus that some people showing PG symptoms were genuinely afflicted by a systematically measurable and sui generis disorder they could not control, then the current equivocation would be a betrayal of patients.

The problem here is not just that we remain somewhat ignorant about PG (and DG). In science we are always somewhat ignorant; but we act, at any given time, as if our most carefully evaluated beliefs are the true ones. The deeper problem, as Shaffer also notes, is that attempts to settle on an operationalization of PG are pulled in three directions. First, there is the fact that “the clinical issue ... is not now, nor has it ever been, whether pathological gambling is an addiction or the result of a biobehavioral vulnerability. From a clinical perspective, this issue involves establishing a working formulation that clinicians and patients can share ... From the treatment side, there is little or no value to understanding any individual as addicted or mentally disordered unless it permits clinicians to choose a treatment plan that will maximize the well-being of the patient.” Second, there is the fact that “[f]or science, the value of improving our understanding of pathological gambling and addiction rests in the development of better theory. Improved theory can guide better research.” Third, and finally, “[f]rom a community perspective, as our understanding of addiction and pathological gambling improve, then the vehicle for more effective social policy emerges.”

Shaffer unintentionally (we presume) exaggerates this tension into being irresolvable by implying that social policy concerns trump the others. “The relationship between pathological gambling and addiction,” he announces, “ultimately rests upon socio-cultural acceptability. After all, “It is best to think of any affliction – a disease, a disability ... – as a text, and of ‘society’ as its author.”⁶⁹ The reason this makes the issue intractable is that ‘society’ would only arrive at a policy consensus if either (i) clinicians developed a single set of diagnostic and treatment practices that were overwhelmingly and demonstratively effective, or (ii) scientific researchers arrived at a clear consensus among themselves on the facts about what PG is – to which ‘society’ could then simply be advised that it had better pay attention, on grounds that preferring falsehoods is just self-defeating.

Now, since PG is clearly a very complex phenomenon, it is hard to imagine any process by which (i) might come about other than by (ii) coming about. Achieving (ii) would not guarantee (i), as the case of cancer shows.⁷⁰ Still, the point remains: either we must hope we can make progress on (ii), and thereby make progress on treatment and policy clarity, or we should give up and admit we’re stuck.

For reasons we’ll present in detail in Chapter 3, we believe that science has recently made considerable progress indeed towards (ii). We stress ‘recently’; the majority of the groundbreaking work we will review has occurred over the past two years. Note that progress towards (ii) does not have to – and in fact has not – consisted in boiling disordered gambling phenomena down to just one thing. Progress can and has consisted in sorting the phenomena clearly into distinct piles.

⁶⁹ Shaffer, H. (1999) Strange bedfellows: a critical view of pathological gambling and addiction. *Addiction* 94: 1445-1448. Shaffer quotes from Blum, A. (1985). The collective representation of affliction: some reflections on disability and disease as social facts. *Theoretical Medicine* 6: 221-232.

⁷⁰ There is general agreement on what cancer is, but for now the syndrome in general remains too complicated to yield univocal treatment approaches even for many specific cancers.

These piles are not importantly sorted, at this point in time, by reference to developmental etiology; we still don't know what causes disordered gambling or any one of its subspecies distinguished in section 1.1, in the sense of knowing why they afflict some people rather than others and in some circumstances rather than others.⁷¹ Instead, the piles are sorted by reference to models of how different disordered gambling syndromes maintain themselves. This is at once useful for both treatment and policy purposes: to the extent that one knows how a syndrome is maintained, one knows, at least in principle, something about what would be required to disrupt that maintenance.

Most of the progress to which we just alluded has been in neuroscience. This is not so surprising: across a wide range of applications, the neurosciences show every sign of being the great scientific story of the 21st century, as physics and information science were in the 20th. Shaffer, notwithstanding his probably unintended undermining of optimism we noted earlier, anticipates this. "An independent gold standard," he wrote (six years ago), "will probably come from neurogenetic or biobehavioral attributes."⁷² We will argue in Chapter 3 that there is now good reason to believe that we have isolated some systematic neurochemical mechanisms that cause a substantial proportion of PGs, as operationalized by DSM-IV, to become incapable of controlling their gambling (at least in the short term) without specific clinical assistance; that we have a broadly correct explanatory model of these mechanisms that enables us to say why it is gambling, and what it is about gambling, that relates these mechanisms to the familiar behavioral patterns traditionally associated with chronic PG; and that at least the broad outline of a standard pharmacological treatment (and perhaps much more than the mere broad outline) is coming into practical view. If we are right about this then, for the first time, it is appropriate to say that there is one kind of disordered gambling – the most severe kind – that we're beginning to truly understand. If we are right about this, then at least that part of the more general DG phenomena should at some point soon be defined in DSM without equivocation. Furthermore, as we will argue in Chapter 4, social policy concerning this part of the phenomenon becomes relatively straightforward and shouldn't even be controversial.

However, we think it is equally clear that the majority of cases where people regret their own gambling behavior are not cases of people in the grip of neurochemically diagnosable pathological gambling.⁷³ In this majority of cases of DG, non-addicted disordered gambling (NADG, which includes the non-addictive cases of what DSM-IV would classify as pathological gambling), all current evidence is compatible with the following conception. All people, as Ainslie argues and as described in sections 1.2 through 1.6, are naturally inclined to discount the future hyperbolically. This means they are all at risk of choosing SSRs over LLRs and then subsequently regretting it. Most people avoid this, most of the time, by finding personal rules that bundle LLRs with choices amongst SSRs. But most people also violate their own personal rules some of the time, and are not usually regarded as therefore having a problem.

⁷¹ We are, though, *starting* to see glimmers of understanding here, as we will indicate in Chapter 3.

⁷² Shaffer, H. (1999) Strange bedfellows: a critical view of pathological gambling and addiction. *Addiction* 94: 1445-1448. p. 1447.

⁷³ This claim is based on US data compiled in Grinols, E. (2004), *Gambling in America*, Cambridge: Cambridge University Press. There are no comparably complete data for South Africa. We will consider these data in Chapter 4, when we come to ask about the implications of successful treatment of different sorts of DG for casino industry revenues.

Indeed, Ainslie argues that healthy people are healthy partly because they break their personal rules from time to time; people who don't do this are tyrants over themselves, and usually viewed by both others and by themselves as rigid and obsessive. But some classes of personal rules tend to get consistently and systematically broken. These fall into two main classes.

First, some people fail to develop effective personal rules in certain whole areas of behavior. Everyone knows otherwise competent people who 'make the same mistakes over and over again' (as we say) in their love lives, or who launch one implausible diet after another. It would be misleading – and almost certainly scientifically incorrect also – to think of these people as addicted to something. But if there are people who are like this with respect to gambling, and if they frequent gambling venues – as they might if that's where their friends go, or if it's the main centre of entertainment in town – then they might well meet the DSM-IV operationalization of pathological gamblers. By the criteria we will present in Chapter 4, they would not be addicts. But they would tend to consistently regret their gambling behavior as their preferences kept reversing following choices of SSRs (in their cases, bets).

Second, there may be sets of environmental circumstances in which almost any personal rule any normal adult might make would tend to be subverted. Most people will blurt out the carefully protected secret, for the momentary pleasure of another's response, when drunk enough – and then subsequently regret it. Most people successfully observe personal rules against adulterous affairs, but most also know that these rules should be applied well before they find themselves alone with temptation in the jacuzzi because at that point it will be too difficult to stop. Casinos are often accused of being places deliberately contrived to make bundling of LLRs with SSRs difficult for most people. Free drinks, jangling coins, burbling and flashing slot machines, interference with the body's natural clock for tracking time of day – all of these are said to be devices that interfere with personal rules. Combined in cunning overall venue designs, they may be sufficient to cause most people to gamble to the point of subsequent regret.

Notice one important difference between these two kinds of case. In the first sort of instance, we still have a distinctive type of person, a person who, as a result of some specific property of theirs, is disposed to gamble problematically. In the second kind of case, we might have problem gambling without there being any problem gamblers, that is, individuals who by dint of special characteristics of their own are more likely to fall into traps. In reality, what is probably most relevant to issues around regulatory policy as applied to casinos is interaction of the two phenomena, in which there are substantial numbers of people who are disposed by personal properties of their own in such a way that their ability to bundle tends to fall apart in casino environments arranged to put pressure on them.

People who are disposed to bundling breakdowns in gambling environments, but who lack the distinctive neurochemical properties we will describe in Chapter 4, will be referred to through the remainder of this report as 'non-addicted disordered gamblers' (NADGs). We do not know of any research that tells us how they are distributed between the problem and pathological gambling classes as DSM-IV criteria would sort this. The reason for this is that the scientific criteria picking out the addicted gamblers (AGs) are still too newly discovered to have featured in any large-scale statistical surveys of DGs. Not, therefore, knowing what proportion of pathological gamblers are AGs, we cannot know how many NADGS (if any) slide all the way to being pathological gamblers. However, in Chapter 4 we will see some reasons for worrying that if a problem gambler becomes a pathological gambler, she may be at risk of being turned, by physical changes in her brain, into an AG.

In any case, we are now in the position we promised to arrive at, of operationalizing problem (and, therefore, disordered) gambling. We will regard as a ‘problem gambler’ anyone who (i) gambles often enough to be consciously aware of trying to develop personal rules to limit their frequency of gambling or their average amount gambled per session or both, but who (ii) regards these rules as inadequately effective because they regret their own gambling behavior on a regular basis, and who (iii) does not meet the DSM-IV criteria for pathological gambling. Then a NADG is any problem gambler or pathological gambler who is not an addictive gambler. A disordered gambler is anyone who is either a NADG or an AG.

Notice that these operationalizations are not made in terms of either raw frequency of gambling or amounts gambled. There are almost certainly people who gamble a great deal, who have no inclination to gamble less frequently or with smaller stakes, and whose brains show no neurochemical traces of addiction. Such people are not, by the criteria given here, DGs. This reflects our BE perspective: the preferences of such people show no reversal, we have no basis for regarding their behavior as irrational or self-defeating, and by hypothesis they have no medical pathology.

This completes the statement of the theoretical framework on which the report is constructed. We use BE analysis to distinguish the overall target of interest, disordered gambling. Later, we will turn to neuroscience – in particular, to a new branch of neuroscience that comports with BE, neuroeconomics, in conjunction with neurochemistry – to isolate a specific subset of disordered gamblers who, we will maintain, warrant a distinctive policy response.

In light of this combination of perspectives, we should mention to the non-scientific reader that it is unusual. Behavioral economists often regard neuroscientists as methodological rivals. This stems from something we described in section 1.2, BE’s emphasis on molar accounts in preference to molecular ones. There are some scientists (‘reductionists’) who think that molecular accounts are the only serious kinds of accounts of phenomena there are. This is dogmatism, not an opinion well supported by the history of science, and behavioral economists are right to resist it. But it would be another kind of (in this case, ‘behaviorist’) dogma to insist that where behavioral patterns are concerned, the best explanation is always and necessarily to be given in molar terms. We think we should let the facts lead methodology rather than the other way around, and that it has just turned out to be the case that most disordered gambling patterns are molar-level phenomena, but that there are AGs whose behavior, in each individual case, is best explained by a recurrent variety of molecular explanation. Even here, our BE perspective still lurks necessarily but invisibly in the background: it is only because the AGs show similar molar behavioral patterns to the NADGs that all are grouped, and usefully so, together as DGs.

We should also mention that we are by no means the first scientific writers on DG to argue there are three types of DGs. They agree with us that there are AGs and NADGs, but then also argue for a third group whose DG behavior is caused by neurochemically and behaviorally identifiable clinical depression. There are indeed data showing correlations between depression and heavy gambling. We have not followed Blaszczynski and Nower’s lead in breaking them out as a distinct category, for methodological reasons related to both scientific and clinical concerns. We are taking our topic to be DG per se. Where AGs are concerned, though a non-trivial proportion of them are multiply addicted, in Chapter 3 we will review evidence that gambling addiction is a sui generis phenomenon onto itself, that produces distinctive patterns of gambling. By contrast, the evidence to this point suggests to us that depression is simply one of the factors that makes some people vulnerable to exploitation as NADGs; depressed people have difficulty bundling.

We expect that the best therapy for most such people is to have their depression treated directly. But as gamblers their patterns seem to resemble those of NADGs rather than AGs, so they do not seem to motivate distinctive gambling-related treatment or policy interventions. We of course remain open to evidence showing that there is either a behaviorally or a neurochemically distinct category of depressed gamblers (distinct, that is, from other DGs and from other depressed people).⁷⁴ If such evidence is forthcoming this will be reflected in future editions of this report.

We conclude this chapter with a summary of the structure of the report to follow. We have now sketched general issues in the scientific study of DG, and have explained the particular conceptual and methodological framework in which we approach it here. Next, in Chapter 2, we will review empirical studies of the behavioral patterns of DGs. As we will see, this work so far leaves considerable gaps in our knowledge. This, it will emerge, amounts mainly to deficits in our knowledge of DG involving NADGs, since we'll show in Chapter 3 that where AGs are concerned the past two years have seen breakthroughs in the achievement of understanding, which point towards development of effective neuropharmacological therapies. Finally, in Chapter Four, we will summarize suggestions for treatment and policy approaches indicated by current scientific knowledge of DG, and also suggest prospective kinds of studies that, in light of current knowledge, look promising as routes to improved treatments and policies.

⁷⁴ Looking ahead to our Chapter 3 discussion of the neuroscience of pathological gambling, or for the benefit of the reader who has studied that section and is now revisiting this one, another basis for our hesitation about 'depressive gambling' is that, unlike the case of AG, there is no identified special neural mechanism for it. Depressed people tend to have low serotonin, and so do AGs. But in the case of AGs this is integrated into a wider neurochemical model, the I-RISA model, and a neuroeconomic model that jointly explain the gambling behavior. In depressed people this is not so. Low serotonin in general (as opposed to the very specific location and dynamics of low serotonin response in AGs' pathways from prefrontal cortex to amygdala caused by neuroadaptation to dopamine surges) is ubiquitous in people with psychological problems. Thus for the moment, we don't know of evidence suggesting that depressed gamblers have distinctive serotonergic patterns that distinguish them from other depressed people.

2. Behavioral Economics, Impulsivity, Addiction and Disordered Gambling

EXECUTIVE SUMMARY

Addicts to and abusers of a wide range of substances are known to be more impulsive than otherwise comparable individuals, in the sense that they have stronger preferences for smaller rewards delivered soon, compared to larger ones that arrive after some delay. Behavioral economists mark this distinction by saying that substance abusers and addicts discount 'more steeply' for delayed rewards.

Another form of discounting concerns not delay, but probability. A person is impulsive with respect to probabilities if their valuation of a reward declines relatively slowly as the likelihood of the reward arriving is reduced. This kind of discounting is less extensively studied than delay discounting, but is clearly of interest for the study of disordered gambling. The empirical studies of delay and probability discounting undertaken to date and including subjects with varying gambling severity (including problem gambling) do not establish clearly whether disordered gamblers are primarily impulsive about delays (like substance abusers) or whether they have distinctive responses to probabilities as well. Further experiments could settle this question relatively simply, and thereby provide a more precise behavioral characterization of disordered gamblers.

Behavioral measures of impulsivity are not the only approaches. That said, behavioral measures, based on experimental measurement, do correlate with gambling severity more successfully than more qualitative and psychiatric measures based on self-report and clinical appraisal. This does suggest that they relate more directly to the phenomena of disordered gambling than psychiatric categories.

Cognitive psychological assessments of decision-making, especially the Iowa Gambling Task, also separate disordered gamblers, who do poorly at them, from normal control subjects. It is not, though, completely clear what the Iowa Gambling Task measures, or how it relates to a number of other important tools for assessing decision-making. Further behavioral experiments could shed light on these questions relatively simply, and thereby enable two relatively isolated bodies of work concerning disordered gambling to be connected. Work in neuroeconomics (surveyed in the following chapter) also suggests ways of connecting the two bodies of work.

Finally, we know that various factors can temporarily influence decision-making, some by changing discount rates. Some individuals may be more vulnerable to these influences, and some environments (such as casinos) might be unusually endowed with them. Both possibilities are relevant to debate over policy regarding gambling and disordered gambling.

2.1. Impulsivity and addiction

In this section we review a range of work showing that behavioral economics has something important to say about addiction in general, especially with reference to the paradigmatic substance addictions including alcohol and nicotine. We go about things this way because it will be useful to establish the credentials of behavioral economics before turning to problem gambling specifically, and also because doing this will establish a framework for discussing some of the peculiarities of gambling compared to substance abuse.

The first sort of work we'll review concerns empirical findings regarding impulsivity and substance abuse. Impulsivity here is to be understood in the narrower behavioural sense, as relatively greater 'steepness' of discounting for delay (see section 1-4 above). There are links between this specific sense of impulsiveness and the more general sense used in psychiatry and personality psychology, but these approaches are discussed separately in connection with work specifically on gambling (section 2-3 below). Some of what is said in the first few paragraphs of this section repeats points made in earlier sections, including parts of section 1-4 and section 1-5. They are included here in the interests of making the present discussion as clear as possible.

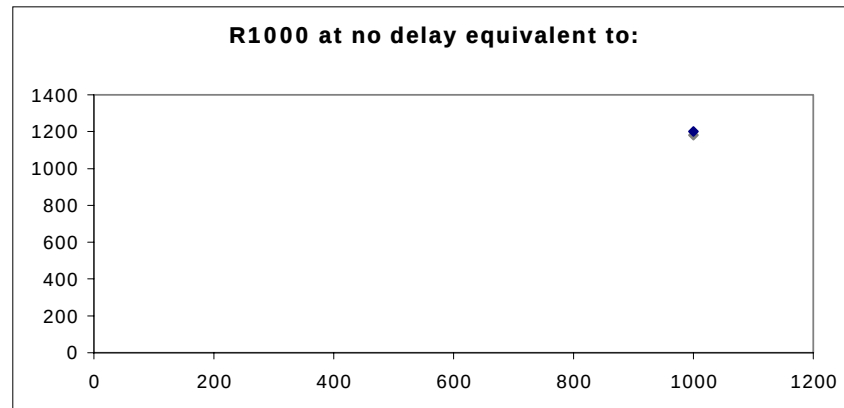
When behavioral economists refer to the 'steepness' of discounting, they have in mind a graphical property of a discount function, itself constructed (for an individual agent) by fitting a curve to a series of indifference points determined by offering a series of choices between pairs of rewards of varying size and delay. The previous sentence contains a lot of jargon, and the next few paragraphs attempt to set out the key ideas referred to, and to outline the experimental procedures used, at least in the case of human subjects.

Anyone concerned to maximise her rewards would rather have R2000 now than R1000 now. Similarly, anyone concerned to maximise her rewards would rather have R1000 now than R1000 after a delay of 1000 days. How about the choice between R1000 now, and R2000 after 1000 days? Most readers of this report should realise that doubling in value in under three years is a very impressive return, one you'd be very unlikely to beat if you took the R1000 and invested it. But most readers also realise that there are some people who would take the R1000 now in this case, and some who would wait for the R2000. That choice splits people into two groups – some preferring the smaller sooner reward, and some the larger later one. What would you do, though, if you wanted a more precise way of distinguishing people with respect to their preferences of this sort? (That is, their preferences as relating to varying delays.)

Suppose that you took a single person, and offered them a whole series of choices, where one option was always R1000 now, and the other was always some amount after 1000 days. Assuming that our subject initially preferred to wait to receive R2000, we would reduce the delayed amount bit by bit, say by R25 every time. At some point we'd find that the subject switched to preferring the smaller sooner reward.⁷⁵ In fact, we'd expect this to happen some time before the later reward comes down to R1000. In between the range of amounts that our subject would not consider worth waiting 1000 days for in preference to R1000 now, and the range that she would consider worth waiting for, is a small range (in theory a single value) where she prefers neither option over the other. In these cases economists would call her 'indifferent' between the options, and the pair of options (amounts and delays) together determine an 'indifference point' that we can plot on a chart of delay versus amount.

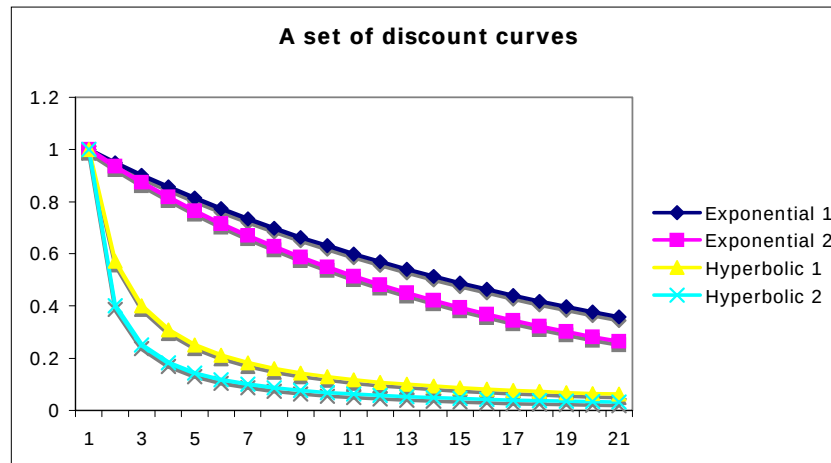
⁷⁵ We're giving, we emphasise, an idealized and simplified image of the experimental procedure here. For a recent example of a real protocol for formulating and delivering the questions to a subject see, for example, Richards, J.B., Zhang, L., Mitchell, S.H., & de Witt, H. (1999) Delay or probability discounting as a model of impulsive behavior: effect of alcohol, *Journal for the Experimental Analysis of Behavior*. 71, 121-143.

That point, to repeat the key idea here, would (in this example) represent an amount of money at a delay of 1000 days that our subject neither preferred nor found inferior to R1000 now. Suppose that the delayed amount in question was R1200, then our point would look like this (with amount on the vertical axis, delay in days on the horizontal one):



We could repeat the process for a number of different delays (1 day, 10 days, 100 days...) and plot the indifference point in each case on the same graph, as long as in each case the 'reference amount' was R1000 now. With several such points, a pattern or definite line may start to get suggested. If the set of indifference points stand in a regular relationship of some sort with each other, then some mathematical equation could describe a line passing through (or very near) those points. Finding such a line is called 'fitting a curve' to a set of points and the resulting equation, in this sort of case, is called a discount function.

The chart below shows value on the vertical axis, and increasing time on the horizontal axis. The lines represent discount curves. What this means is that each point on a curve represents the value of a reward at a certain delay, and the collection of points represents the way in which the value of a reward declines as the delay before its delivery increases. The topmost two lines are drawn using an 'exponential' discount curve, the other two are drawn using 'hyperbolic' discount curves. The two hyperbolic curves are drawn using the same equation, but with a different value for the discount rate in each case. When we say that some person, or type of person, discounts 'more steeply' than another, we mean that their discount rate is higher, and hence that the value of a reward declines more rapidly as delay to delivery increases. The 'steepness' is the steepness of the downward slope from the point of immediate delivery.



In the curve fitting process, the vast majority of the experimental data concerning humans and animals finds that delay discounting fits a hyperbolic curve better than an exponential one. (See section 1-5 above) The steepness of an individual discount curve or function is represented by the value of a single parameter in the equation. An example of a simple hyperbolic formula fitting most of the data very well is equation (3) discussed in section 1-4 above.⁷⁶

$$v_i = \frac{A_i}{1 + kD_i} \quad (3)$$

In this formula, recall, v_i , A_i , and D_i represent the present value of a delayed reward, the amount of a delayed reward, and the delay of the reward, respectively, and k is a constant representing degree of impatience. In the graph above the steeper of the two hyperbolic curves has a ' k – value' of 1.5, the shallower one a k – value of 0.75. Most claims about the relative steepness of discounting in a comparison of two individuals or groups are claims about different values or average values of the k parameter manipulated in the curve-fitting process.

The results we'll report almost all take the form of associations, or correlations, to the effect that in some or other population of substance abusers or addicts, delay discounting was found to be steeper than for control subjects, or steeper than some norm or other contrast class. That is, that addicts and substance abusers were less tolerant of waiting, or were more impulsive, in preferring a smaller reward sooner than a larger later one to a greater degree, than non-addicts and abusers.

Such correlations do not by themselves establish causation. That is, they don't tell us by themselves whether the individuals in question became abusers because they were disposed to be more impulsive beforehand, or whether use (or abuse) of a substance had made them more impatient.

⁷⁶ Mazur, J.E. (2001). Hyperbolic value addition and general models of animal choice. *Psychological Review*, 108, 96-112. For a review of empirical results and the use of different formula (not significantly different for present purposes) to account for the results, see: Green, L. & Myerson, J. (2004) A Discounting Framework for Choice With Delayed and Probabilistic Rewards, *Psychological Bulletin*, 130(5), 769-792.

It could well be that both are true – that some people are prone to immoderate substance use because of being a bit less patient, and that this consumption makes them even more impulsive, so that they become ‘proper addicts’ with time. (For a more thorough discussion of the difference between real addicts and other sorts of disordered consumption see Chapter 3 below.)

To shed light on causation we’d need more ‘longitudinal’ data than we have.⁷⁷ That is, we’d need measurements of the discount rates of whole cohorts of people over a number of years from when they were relatively young, so that follow up study could establish the extent to which this or that degree of impulsiveness in youth was a predictor of a self destructive pattern of behavior, including disordered gambling, in later life. Fortunately there are some suggestive studies with a longitudinal component, indicating that the effort could be worth making. In one such study impulsivity at age 12-14 was found to be a significant predictor of involvement in problem gambling at age 17.⁷⁸ That is, younger boys who disproportionately (compared to their peers) preferred smaller sooner rewards to larger later ones, were more likely than their peers to be disordered gamblers a few years later. A related finding is that low impulsivity is a good predictor of academic success.⁷⁹ This isn’t surprising, since the rewards of academic study are delayed, so impulsivity can be expected to manifest in the costs of present effort not seeming worth the distant rewards.

These last two results are also of interest since it suggests that a ‘common cause’ may explain both some of the non-gambling problems in the lives of disordered gamblers, and help explain their disordered gambling, rather than most of the problems being the consequence of the gambling. That is, in at least some subjects it seems plausible to suppose that their failure to complete post-secondary qualifications, or keep up payments on investments, or otherwise follow through on longer-run plans might be the product of the same underlying impulsivity as their disordered gambling, rather than being casualties of the need to feed their condition. We return to these and related questions in our discussion of policy implications in chapter 4.

Other useful work compares the discount rates of current, former, and never consumers of some addictive substance, and finds, for example, that matched ex-smokers and never smokers both discount less steeply than current smokers, suggesting either that smokers with low discount rates are more likely to become ex-smokers or that stopping smoking is associated with a change in discounting behavior. Note that this work doesn’t directly tell us what causes what, since the data in this study is not truly longitudinal and suggests only that both smoking status and discounting change at more or less the same time.

⁷⁷ This isn’t all we’d need. Existing work on brain development, and ongoing work on how the brain implements choice-making, will also play a part in figuring out what is going on here. See chapter 3 below for a review of neuroeconomics.

⁷⁸ Vitaro, F., Arseneault, L. & Tremblay, R. E. (1999) Impulsivity predicts problem gambling in low SES adolescent males, *Addiction*, 94(4), 565-575. Note that the measure of impulsivity here was not behavioral but clinical. Since other work has found a behavioral measure (discounting task) to be a better predictor of gambling severity within a population at a time (Alessi, S.M. & Petry, N.M. (2003) Pathological gambling severity is associated with impulsivity in a delay discounting procedure, *Behavioural Processes*, 64, 345–354.), it may be that a stronger result would have been found with a behavioral measure.

⁷⁹ Kirby, K. N., Winston, G. C. and Santiesteban, M. (2005). Impatience and grades: Delay-discount rates correlate negatively with college GPA. *Learning and Individual Differences*, 15, 213–222.

It does, though, point to the possibility of changes, including reversible changes, in discount rates over extended periods,⁸⁰ and suggests a search for the determinants of such changes.

Finally, in the laboratory, early impulsivity is a predictor of the extent of cocaine self-administration in rats.⁸¹ Manipulation of reward-relevant chemical activity in the brain is followed by changes in the steepness of discounting in rats, providing further encouragement to those who seek a drug-based therapy for at least some problem gambling.⁸² (See chapter 3 below for much more on this topic.)

Although associations shed little direct light on causation, the ones we shortly review are informative and useful. This is for several reasons. First, that discounting tasks can find discount economic approach to studying addiction. Unless such people were able to make reasonably regular comparisons of the relative values of rewards of different magnitudes and at different times, it would not be possible to measure their discount rates with any confidence. Second, the fact that addicts tend to discount more steeply is reason to suppose that the behavioral notion of impulsivity, as measured by relative discount rates, is a good tool for sorting disordered consumers from normal ones, and a good basis for trying to work out what it is that makes the difference. After reviewing some of the evidence that behavioral notions of impulsivity apply to addictions, we note some pieces of evidence bearing on the interpretation of the associations we report.

One study compared regular smokers with non-smokers on a variety of clinical and personality measures of impulsivity, as well as three distinct behavioral tasks, assessing impulsivity as a function of delay, as a function of reduced probability and as a function of effort required to obtain a reward. It found that smokers had higher impulsivity scores than most controls on the personality questionnaires, and that they were more impulsive when it came to delays, but not on the other behavioral tasks.⁸³

A complementary study of current smokers, former smokers, and people who had never smoked at all found that current smokers discounted the value of delayed money more steeply than both never smokers and ex smokers. Smokers also discounted the value of delayed cigarettes more steeply than they did money. Never and former smokers were not significantly different from each other. This work suggests, but does not show, that smoking is associated with a temporary and reversible change in discounting.⁸⁴

⁸⁰ Older subjects appear to discount the future less steeply than younger ones. See Green, L., Fry, A. and Myerson, J., 'Discounting of Delayed Rewards: A Life-span Comparison', *Psychological Science* 5 (1994): 33–36.

⁸¹ Perry, J.L., Larson, E.B., German, J.P., Madden, G.J., & Carroll, M.E. (2005) Impulsivity (delay discounting) as a predictor of acquisition of IV cocaine self-administration in female rats, *Psychopharmacology*, 178(2 – 3), 193 – 201.

⁸² Bizot, J-C., Le Bihan, C., Puech, A.J., Hamon, M. & Thiébot, M-H. (1999) Serotonin and tolerance to delay of reward in rats, *Psychopharmacology*, 146: 400–412.

⁸³ Mitchell, S.H. (1999) Measures of impulsivity in cigarette smokers and non-smokers, *Psychopharmacology*, 146, 455-464. The mean age of subjects in this study was a little over 20 years.

⁸⁴ Bickel, W., Odum, A. & Madden, G. (1999) Impulsivity and cigarette smoking: delay discounting in current, never, and ex-smokers, *Psychopharmacology*, 146, 447-454.

The result that current smokers discount more steeply than non-smokers has been confirmed in other research.⁸⁵

In one of the first studies of this sort, the discount rates of a mixed population of inpatient substance abusers was compared to the discount rates of those who worked at the unit, and it was found that the substance abusers discounted delayed rewards more steeply.⁸⁶

Considering alcohol specifically, it has been found that college students who were heavy and problem drinkers discounted for delay more steeply than light drinkers.⁸⁷ A further study comparing currently drinking alcoholics, abstaining alcoholics, and nonalcoholic control subjects found that drinking alcoholics discounted more steeply than abstainers, who in turn discounted more steeply than non-alcoholics. Alcoholic subjects discounted alcohol more steeply than money.⁸⁸ Finally, it was found in a further study that discount rates were associated (in a further sample of college students) with a number of substance consumption variables, including age of first use of alcohol, tobacco and marijuana, number of times incapacitated by alcohol, and total number of illegal drugs used.⁸⁹

Turning to drugs other than alcohol, in a pair of careful studies the discount rates of opioid dependent subjects recruited from a treatment programme were compared to those of control subjects matched for age, education, gender, IQ and income, and drawn from the same community. Opioid dependent subjects discounted the value of delayed hypothetical money substantially more steeply than did controls. In addition, they discounted the value of delayed hypothetical heroin even more steeply than they did money.⁹⁰ These studies, and the second of the studies on discounting and alcohol described above, both show that there are subject specific variations in the rate at which different sorts of reward are discounted, and that discount rates can be significantly higher for rewards of abuse.

⁸⁵ Reynolds, B., Richards, J. B., Horn, K. and Karraker, K. (2004). Delay discounting and probability discounting as related to cigarette smoking status in adults. *Behavioural Processes*, 65, 35–42. This study also examined discounting for reduced probability – see section (2-2) below for a discussion of this form of discounting.

⁸⁶ Ainslie, G. & Haendel, V. (1983) The motives of the will, in: Gottheil, E., Druley, K., Skodola, T. & Waxman, H. (Eds) *Etiology Aspects of Alcohol and Drug Abuse*, pp. 119-140 (Springfield, IL, Charles C. Thomas).

⁸⁷ Vuchinich, R. E. & Simpson, C. A. (1998) Hyperbolic temporal discounting in social drinkers and problem drinkers, *Experimental and Clinical Psychopharmacology*, 6, 292-194.

⁸⁸ Petry, N.M. (2001) Delay discounting of money and alcohol in actively using alcoholics, currently abstinent alcoholics, and controls. *Psychopharmacology*, 154(3), 243-250.

⁸⁹ Kollins, S. H. (2003). Delay discounting is associated with substance use in college students. *Addictive Behaviors*, 28, 1167–1173.

⁹⁰ Madden, G.J., Bickel, W.K. & Jacobs, E. A (1999) Discounting of delayed rewards in opioid-dependent outpatients: exponential or hyperbolic discount functions, *Journal of Experimental and Clinical Psychopharmacology*, 7, 284-293. And, Madden, G.J., Petry, N., Badger, G. & Bickel, W.K. (1997) Impulsive and self-control choices in opioid-dependent subjects and non-drug using controls: drug and monetary rewards, *Experimental and Clinical Psychopharmacology*, 5, 256-262.

A further study using real (i.e. non-hypothetical) money also found that heroin-using subjects discounted dramatically more steeply than non-heroin using controls, who had again been matched carefully in other respects,⁹¹ and that the same went for cocaine-abusers.⁹² A complementary study found that among a population of heroin users, all of whom discounted very steeply, those who discounted most steeply were also those most likely (on the basis of self-report) to share a needle in a case where sufficient sterile ones were not available.⁹³ It has also been found that mild deprivation from opioid drugs leads to steeper discounting of delayed rewards in already dependent subjects.⁹⁴

This collection of results certainly encourages the view that steepness of delay discounting is an important factor distinguishing a variety of substance abusers from non-abusing subjects. At the same time this therefore suggests that the general behavioral economic framework for the study of decisions is appropriate for studying the decisions of addicts, or at least substance addicts. What about disordered gamblers?

2.2. Discounting: delay and probability

The great majority of the empirical work concerning discounting in human subjects, including work on addictions, focuses, as was explained in section (1.4) above, on discounting as a function of delay. Being removed in time, though, is not the only way in which a reward of a certain magnitude can be separated from consumption. Nor, for anyone interested in gambling, is it the most interesting. The other main way in which the value of a reward of a certain magnitude can be manipulated is by varying the probability that it will arrive.⁹⁵ That is, just as the present value of R100 after a delay of 1 month is, for most agents, lower than that of R100 right now, so the value of a 50/50 chance of getting R100 is, again for most, less than that of a guaranteed R100. Since gambling essentially involves risk and uncertainty, it is plausible to suppose that gamblers, and especially disordered gamblers, might be distinguished specifically by different responses to risk than control subjects. Disordered gamblers, rather than merely having different responses to delay in common with other addicts, might be specifically unusual in their responses to risk, and may even be no different to non-disordered gamblers in their responses to delay.

⁹¹ Kirby K.N., Petry, N.M. & Bickel, W.K. (1999) Heroin addicts discount delayed rewards at higher rates than non-drug using controls. *Journal of Experimental Psychology: General Processes*, 128:78–87.

⁹² Kirby, K. N. & Petry, N. M. (2004) Heroin and cocaine abusers have higher discount rates for delayed rewards than alcoholics or non-drug-using controls, *Addiction*, 99, 461–471.

⁹³ Odum, A. L., Madden, G. J., Badger, G. J. & Bickel, W.K. (2000) Needle sharing in opioid-dependent outpatients: psychological processes underlying risk, *Drug and Alcohol Dependence*, 60, 259– 266.

⁹⁴ Giordano, L. A., Bickel, W. K., Loewenstein, G., Jacobs, E. A., Marsch, L. and Badger, G. J. (2002). Mild opioid deprivation increases the degree that opioid-dependent outpatients discount delayed heroin and money. *Psychopharmacology*, 163,174–182.

⁹⁵ For a thorough review of the literature on delay and probability discounting see Green, L. and Myerson, J. (2004). A Discounting Framework for Choice With Delayed and Probabilistic Rewards. *Psychological Bulletin*, 130(5), 769–792.

Whereas, in the case of delay, the discounting of an individual subject is assessed by asking a series of questions about which of two rewards, varying in size and delay, she prefers (e.g. would you rather have R100 now, or R120 in a week?), in the case of probability, the questions relate pairs of amounts and chances (e.g. would you rather have R100 now, or a 50/50 chance of R250?). As in the case of delay discounting, a discount function for an individual agent is constructed by fitting a curve to a series of indifference points determined by offering a series of choices between pairs of rewards of varying size and likelihood. And, as in the case of delay discounting, the curve that best fits most of the experimental data on the patterns of choices of real agents, animal or human, is a hyperbolic one. The steepness of this curve is a measure of risk sensitivity, just as the steepness of the delay discount curve is a measure of impulsiveness.

Part of the reason that we know so little is that delay discounting promises a general explanation of self-defeating behaviour, because (as explained in Chapter 1) the fact of hyperbolic discounting for delay predicts that agents' priorities will not be stable over time, and hence that people will make plans only to break them, forswear temptations only to give into them, and regret these patterns of behaviour. Not only that, some theorists have held that probability discounting is a derivative form of delay discounting, and hence of limited independent interest.

Part of why they thought this was that it seemed likely to them that the two sorts of discounting distinguished here are manifestations of a single, fundamental, discounting process – either concerning chances, delays or something not quite either. Among the reasons for thinking this are that any delay to delivery of a reward involves an element of risk – especially risk that you won't be there still to collect it, that some misadventure including the efforts of a competitor will take it away between now and scheduled delivery. Thus delay and uncertainty are connected in everyday contexts.⁹⁶ Not only that, some argue that the fact that the same (hyperbolic) form of mathematical function describes both sorts of discounting data should lead us to think that there is a single underlying process. There is also some evidence that in individual subjects the rate of discounting for delay and that for probability are correlated.⁹⁷ Unfortunately, though, there have been relatively few studies in which the delay and probability discounting of subjects have both been measured and compared, so we know too little about the extent to which they typically relate to one another. Given the importance of risk to gambling, considerably more study on probability discounting (in populations with mixed gambling severity) is required.

⁹⁶ This view is common in the ethological (animal behaviour) literature where delay standardly takes place in environments with competitors, who can help themselves to your unexercised options. See, for example, Houston, A., Kacelnik, A., & McNamara, J. (1982). Some learning rules for acquiring information. In D. McFarland (Ed.), *Functional ontogeny* (pp. 140–191). Boston: Pitman; Stephens, D. W., & Anderson, D. (2001). The adaptive value of preference for immediacy: when short-sighted rules have far-sighted consequences, *Behavioral Ecology* 12:330-339; Stephens, D. W., Kerr, B. & Fernandez-Juricic, E. (2004) Impulsiveness without discounting: the ecological rationality hypothesis, *Proceedings of the Royal Society of London (B)* 271: 2459-2465. A version of this view is also defended in Rachlin, H., Raineri, A., & Cross, D. (1991). Subjective probability and delay. *Journal of the Experimental Analysis of Behavior*, 55, 233–244.

⁹⁷ Richards, J.B., Zhang, L., Mitchell, S.H., & de Witt, H. (1999) Delay or probability discounting as a model of impulsive behavior: effect of alcohol, *J. Exp. Anal. Behav.* 71, 121-143. A weak positive correlation between the two measures is also reported by Mitchell, S.H. (1999) Measures of impulsivity in cigarette smokers and non-smokers, *Psychopharmacology*, 146, 455-464.

Against the suggestions that the two kinds of discounting might be manifestations of a single process, there are several ways in which the two apparently come apart ('dissociate', as scientists say). One study asked subjects about the relative preferability of pairs of immediate and delayed rewards in each of two different currencies. One of the currencies (the Polish Zloty) was the money of a country undergoing rapid inflation, while the other (the American Dollar) was not. The experimenters found that subjects treated the chances of immediate rewards of the same magnitude equivalently irrespective of currency. (That is, their discounting for reduced likelihood was the same irrespective of currency.) However, when considering guaranteed delayed rewards, they discounted the rapidly inflating currency more steeply.⁹⁸ This selective effect on one kind of discounting tells against the view that there is a single discounting process. In a complementary study it was found that subjects' cultural background could make a selective difference to one or the other kind of discounting.⁹⁹ In other work it was found that while smokers discounted more steeply for delayed rewards than non-smokers, and were less sensitive to reduced probabilities, the difference between smokers and non-smokers was much more pronounced in the case of delay discounting.¹⁰⁰

More importantly for present purposes, there is evidence that some gamblers differ from non-gamblers (understood as people with little or no experience of wagering or otherwise paying to participate in games of chance) in their probability discounting, but not their delay discounting. In one study at least, it has been found that subjects that gambled (and who scored in the 'problem' range on a standard gambling severity measure) were less sensitive to reduced probability than non-gamblers, but did not discount delayed rewards more steeply than non-gamblers.¹⁰¹ This result is apparently in conflict with the finding of another study to the effect that pathological gamblers did discount more steeply for delay than non-problem controls (and, consonant with prior work on abusers of more than one substance, that pathological gamblers with substance abuse problems discounted yet more steeply for delay).¹⁰²

⁹⁸ Ostaszewski, P., Green, L. & Myerson, J. (1998) Effects of inflation on the subjective value of delayed and probabilistic rewards. *Psychonomic Bulletin & Review* 5:324–33.

⁹⁹ Du, W., Green, L., & Myerson, J. (2002). Cross-cultural comparisons of discounting delayed and probabilistic rewards. *Psychological Record*, 52, 479–492. In this study the discounting of American, Chinese, and Japanese graduate students was assessed. The Chinese discounted delayed rewards more steeply than the Japanese, whereas the Japanese discounted probabilistic rewards more steeply than the Chinese. (The Americans discounted relatively steeply in both areas)

¹⁰⁰ Reynolds, B. Richards, J.B., Horn, K., & Karraker, K. (2004) Delay discounting and probability discounting as related to cigarette smoking status in adults, *Behavioral Processes*, 65, 35–42. The mean age of subjects in this study was 20 years. Another study found a smaller effect for delays, and none for probabilities: Mitchell, S.H. (1999) Measures of impulsivity in cigarette smokers and non-smokers, *Psychopharmacology*, 146, 455–464.

¹⁰¹ Holt, D. D., Green, L. and Myerson, J. (2003). Is discounting impulsive? Evidence from temporal and probability discounting in gambling and non-gambling college students. *Behavioural Processes*, 64, 355–367. The mean age of subjects in this study was 19.6 years. Note that the gamblers had South Oaks Gambling Screen (SOGS) scores of 4 or more, whereas control subjects included persons with SOGS scores of zero or 1.

¹⁰² Petry, N. M. (2001) Pathological gamblers, with and without substance abuse disorders, discount delayed rewards at high rates, *Journal of Abnormal Psychology*, 110 (3), 482–487. The mean ages of participants in this study were 39 for controls, 44 for non-substance abusing pathological gamblers, and 43 for substance using pathological gamblers.

This doesn't suggest that gambling is a substance, but rather that disordered gambling seems to be behaviorally similar in a key way to substance addictions – having more than one consumption problem, even if only one is to a substance, seems to be associated with less severe discounting than when you have two of them.

Direct comparison of the two studies is difficult given that the gamblers in the second study (finding an effect in the case of delay) had much more severe gambling problems (with SOGS scores of around 12-13, compared to 6-7) than those in the first study (finding an effect for probability but not delay). Subjects in the second study were also older (with a mean age of 44, compared to 19.6) and with an average of 8 years of gambling problems each. The explanation for this difference is clear enough: one study used students as subjects, the other recruited subjects from a problem gambling treatment facility. This contrast and the possible conflicting results of the two studies, illustrates the need for larger studies in which as far as possible matched subject populations nonetheless differ significantly in gambling severity. Further explanation for the discrepancy is given in Chapter 3 below, in the discussion of neuroeconomic findings about disordered gambling.

Not only are delay and probability dissociable, but most real choices, pretty obviously, involve elements of both kinds of separation. That is, most real choices concern rewards that won't arrive immediately, and which are also not absolutely guaranteed to arrive. In particular, when people gamble, they explicitly make such choices, declaring by betting that they'd rather have the chance of having whatever the bet could win, when the outcome is decided or determined, than have some of the money they already have, and that they pay to place the bet. Since gambling involves uncertainty, it would be useful to know a lot more about probability discounting, and especially population differences in its expression, than we presently do.

Among the things it would be useful to know is how exactly non-disordered gamblers, and disordered gamblers of differing severity of problem, differ. Perhaps, for example, moderately disordered gamblers are more like substance abusers in being relatively more sensitive to delay, but don't differ in their probability discounting, whereas severely disordered gamblers are both impatient and insensitive to risk. If there are indeed subjects (and severely disordered gamblers are a likely example of this) who are impulsive and risk-insensitive, then we need to conduct research on how particular situations, and particular series of choices and results of choices, might be especially destructive for such people. Casinos, for example, may turn out to be exquisitely tailored traps for such people. Or it may turn out that sufficiently extreme discounting makes the whole world a trap.

Before saying more about what we could find out about disordered gambling with more and better research on discounting, though, a look at some of the existing results concerning disordered gambling is in order.

2.3. Impulsivity and Disordered Gamblers

The discussion of behavioural economics, and its results with reference to addicts of various sorts, over the preceding pages might have given the impression that there is a single science of behaviour, where even though there is not full agreement about how to explain the data, all practitioners are working with one set of empirical instruments for gathering the data itself. This impression would be misleading. Alongside, and for historical reasons significantly independent of, the tradition of behavioural research are other streams of enquiry with their own methods and tools. Among the most important of these for the purposes of the study of disordered gambling are cognitive psychology and psychiatry.

Behavioural research, as we have seen, emphasises the economic dimension of choice, considering the patterns of costs and benefits of actions, the budget constraints of the agents that make them, and so forth. The tools of behavioural research are experimental protocols for extracting the data required to construct mathematical representations of aspects of the choice making of an individual agent. An example of this is the series of options presented so as to determine a set of 'indifference points' (see section 1-5 and 2-1 above), through which a curve representing a discount function can be drawn. To recap, an indifference point represents a pair of options (values and delays, or values and probabilities) of which a given agent finds neither preferable to the other. A set of such points together determine a 'discount curve' which is a profile of one aspect of an agent's choice behaviour, for example sensitivity to delay (see the graph in section 2-1 above). Differences between individuals and groups can sometimes be described as differences in the properties of these curves. We saw examples of this above where substance addicts of various sorts were found to discount delays more steeply than non-addicts.

Cognitive psychology for its part understands decision-making as a process primarily focused on the acquisition and manipulation of knowledge and the control of action, largely independently of questions about what makes some outcomes desirable, or what motivates agents at all. The historical reasons for the relative neglect of motivation by cognitive psychology aren't relevant to this survey, but the fact of it helps explain why cognitive psychologists have studied various kinds of decision-making, including pathological patterns of choosing such as addictions, in different ways. Cognitive psychologists tend to be more concerned than behaviourists about the architecture of the processing systems (the brain and its sub-systems) involved in decision, and in how the inputs to the decision process are represented and manipulated. The new field of neuroeconomics, combining tools and techniques from behavioural economics and cognitive neuroscience, is a particularly clear indication that long-overdue rapprochement is underway. This field is discussed in Chapter 3 below.

Psychiatry, finally, has considered decision, especially pathological decisions, in a way intended to be continuous with the rest of medical science. Psychiatry is often defined as a branch of medicine, and unlike behavioural economics and cognitive psychology, its subject matter is in turn often defined in terms of mental illness, and its treatment and prevention. Much psychiatry as relevant to disordered gambling draws on the field of personality psychology, attempting to develop frameworks for classifying all humans by analysing personality into some or other set of dimensions (such as 'sociability').¹⁰³ Most theories of personality attempt to classify subjects' impulsivity, but there is considerable variation in whether impulsivity is a dimension itself (so a basic fact about a person is his or her level of impulsivity) or whether impulsivity is a name for a more complex trait depending on how a subject is ranked on two or more other dimensions. The lack of consensus within psychiatry and personality psychology about the measures of impulsivity suggests that all is not well with the measures and the theories that inform them. This is especially so given the precision and power of behavioural approaches, meaning that there is an attractive alternative. Some comparative research suggests that disordered gamblers are classified correctly with psychiatric impulsivity measures, but with less effective discrimination than behavioural ones.¹⁰⁴

¹⁰³ Older approaches in psychology distinguished 'psychoses' (involving more serious symptoms such as delusions) and less severe 'neuroses', as well as categories such as 'dementias'.

¹⁰⁴ Alessi, S.M. & Petry, N.M. (2003) Pathological gambling severity is associated with impulsivity in a delay discounting procedure, *Behavioural Processes*, 64, 345–354. This paper found that differences in delay discounting predicted South Oaks Gambling Screen scores 1.4 times as well as scores on the Eysenck impulsivity scale.

Given this, it is not surprising that various studies report that disordered gamblers are more impulsive by the lights of personality psychology measures of impulsivity,¹⁰⁵ or that high scores on these measures at one age is a predictor of disordered gambling later on.¹⁰⁶

All three of these main approaches (behavioural economics, cognitive psychology, psychiatry) have an established interest in disordered gambling, and all have something to say about how disordered gamblers differ from other people, and about what might make the difference. There is, though, no simple fact of the matter about how the findings of the three approaches introduced here relate to one another. In some cases, as we explain, they say complementary things, in others they seem to conflict, while in some it is hard to tell whether they are talking about the same thing at all. This is because the operational content of their different tools and notions of impulsivity differ in various ways, and there is very little experimental work aimed at determining how the different measures relate to one another in a single population.

In case this fragmented image of the sciences of human choice seems far-fetched, bear in mind two things. First, the phenomena are very rich and complex. They comprise, after all, human patterns of choice in varying environments, and the responses of other humans to the choices made by one another. Second, divisions within science can persist for extended periods if supported by divisions in how practitioners are trained. The position of psychiatry within the medical establishment, while other approaches are largely outside it, is one example of this. The further division within scientific psychology between cognitively oriented researchers and behaviourally minded ones is another. As we've mentioned before, and will soon demonstrate at length, the rising science of neuroeconomics is an illustration that a new generation of practitioners who bridge cognitive and behavioural approaches are on the rise.

One widely used experimental task in the cognitive science literature is the 'Iowa Gambling Task' (IGT).¹⁰⁷ Since there is a body of work finding that performance on this task separates addicts of various kinds, including gambling addicts, into one pile and normal subjects into another, there is reason both to want to know how and why it works, and also to know how it might relate to the discounting procedures that we've already seen sorting the subjects into the right piles.

In the Iowa Gambling Task¹⁰⁸ a subject is asked to choose a series of cards from four decks. The cards are face down, so subjects are uncertain about the consequences of their choices until each one is made, and the selected card turned over. The subject can choose a card

¹⁰⁵ See the following four studies: Steel, Z. and Blaszczynski, A. (1998). Impulsivity, personality disorders and pathological gambling severity. *Addiction*, 93(6), 895-905. Blaszczynski, A., Steel, Z. & McConaghy, N. (1997) Impulsivity in pathological gambling: the antisocial impulsivist, *Addiction*, 92(1), 75-87. Blaszczynski, A. & Nower, L. (2002) A pathways model of problem and pathological gambling, *Addiction*, 97, 487-499. Petry, N. M. (2001). Substance abuse, pathological gambling, and impulsiveness. *Drug and Alcohol Dependence*, 63, 29-38.

¹⁰⁶ Vitaro, F., Arseneault, L. and Richard E. Tremblay, R. E. (1999). Impulsivity predicts problem gambling in low SES adolescent males. *Addiction*, 94(4), 565-575.

¹⁰⁷ It is named, like a number of other tests and experimental tasks for the university at which it was developed.

¹⁰⁸ Bechara, A., Tranel, D., & Damasio, H. (2000). Characterization of the decision-making impairment of patients with bilateral lesions of the ventromedial prefrontal cortex. *Brain*, 123, 2189-2202.

from any deck at any time, and are told to try to maximise their earnings over the course of the task. Each card leads to some combination of winnings and penalties, for example winning R250, and paying R50. Unknown to the subjects, two of the four decks are 'bad' decks in which large wins are on average more than cancelled out by regular penalties, so that consistently choosing from those decks is guaranteed to lose money. The two other decks are 'good' ones where the balance of reward and penalties favours the subject who persists with those decks, even though the rewards associated with individual cards are generally smaller. Although subjects aren't told this, a standard implementation of the test involves turning 100 cards. The performance of a subject is determined by counting the changing proportion of choices from good and bad decks made as the subject is exposed to more and more cases of the effects of choosing from each deck. Most subjects start allocating their choices roughly equally, and a 'normal' or 'healthy' subject should learn to make fewer and fewer selections from the bad decks as they proceed through the task.

Some famous early research using the task found that patients with damage to the ventro-medial prefrontal cortex performed much worse than undamaged controls.¹⁰⁹ In light of other results with the same subjects and certain other considerations, it was proposed that normal subjects successfully formed unconscious¹¹⁰ 'somatic markers' or emotional associations with actions and the consequences of choices and that disruption of this process could lead to impaired decision making.

From the perspective of behavioural economics, the Iowa Gambling Task is, it must be said, a rather peculiar research instrument. There is no good reason to have four decks rather than between pairs of options, behavioral experiments have largely focussed on tasks and situations with only pairs, to keep things no more complicated than they need to be. (Indeed a variation of the task developed for use with children dispenses with half of the decks for the very reason that the extra two aren't required, or even distinguished in scoring the performance of subjects.¹¹¹) There's also enough that the subjects don't know to confound interpretation of some of their behaviour. For a start they don't know for sure what the gains from any deck will be, even though those are fixed for each deck. This means that, unlike a discounting task of the sort described in the previous section, choices don't directly convey information about the individual subject's response to risk, even though the risk of gains and punishment is clearly important in the task. As subjects proceed with the task they do get exposed to information about the frequency and magnitude of penalties in each deck, but don't ever know those contingencies in advance of their choices either. Again, interpretation of the results cannot clearly say whether the behaviour of subjects is telling us something about their preferences, for example that they favour the bad decks just as they favour all risky options, or their ignorance, in case they have failed to come to understand that those decks are bad at all.

¹⁰⁹ Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50, 7–15.

¹¹⁰ There is some controversy over the extent to which the 'somatic marker' is unconscious. This issue does not make a difference to the issues discussed in the present survey. See Maia, T.V. & McClelland, J.L. (2004) A reexamination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa gambling task, *Proceedings of the National Academy of Sciences*, 101 (45), 16075–16080. For a reply see Bechara, A., Damasio, H., Tranel, D. & Damasio, A.R. (2005) The Iowa Gambling Task and the somatic marker hypothesis: some questions and answers, *Trends in Cognitive Sciences*, 9(4), 159–162.

¹¹¹ Kerr, A. and Zelazo, P. D. (2004). Development of "hot" executive function: The children's gambling task. *Brain and Cognition*, 55, 148–157.

(In fact, there is a variation on the Iowa task that makes the likelihoods known up front for that reason.¹¹²) No matter what they know, their choices involve a considerable mixture of delayed and probabilistic rewards and penalties. That is, if they assume (probably tacitly) that the remainder of each deck is pretty much like what they have experienced it to be, then their view of that deck will include a mixture of frequencies and probabilities of gains and losses of various sizes in a way that leaves it quite unclear how their choices relate to their individual discount curves for delay and probability. The number of choices made by subjects is also too small to disentangle these components. In a behavioural economics discounting task, a human subject will typically have to answer around 70-90 questions in order to locate enough indifference points to determine a discount curve with reasonable confidence. To measure both delay and probability discounting for a single subject accurately, then, would take nearly twice as many questions as one run of the Iowa Gambling Task. Yet discounting is surely a factor in performance, for reasons we return to shortly.

More recent work has suggested that some of what the Iowa Gambling Task has purportedly measured may not be what it was intended to measure. A team of researchers developed a variant of the task, called the Soochow Gambling Task, in which the 'good' and 'bad' decks had the same average net earnings as in the Iowa Task, but where the frequency of rewards was different. This research found that normal subjects (that is, subjects who were not brain-damaged in areas known to correlate with bad Iowa Task performance) who did 'well' at the Iowa Task, made mostly bad choices in the Soochow Task when the frequency of reward from the bad decks was high. On the basis of this work some of the early research on the Iowa Task has been reinterpreted, suggesting that frequency of reward is a much more important determinant of how normal subjects behave.¹¹³

Having noted these concerns and difficulties, it should also be said that some significant results have been found with the Iowa Gambling Task as well. Although it has not been extensively used as a tool for the study of addiction, it has been found that subjects with substance dependence (the dependencies in the study being referred to here were of various sorts including alcohol, crack-cocaine and methamphetamine) have been found to perform poorly at the IGT compared to normal controls.¹¹⁴

¹¹² Brand, M., Kalbe, E., Labudda, K., Fujiwara, E., Kessler, J., & Markowitsch, H.J. (2005) Decision-making impairments in patients with pathological gambling, *Psychiatry Research*, 133, 91– 99.

¹¹³ Chiu, Y-C., Lin, C-H., Huang, J-T., Lin, S., Lee, P-L, & Hsie, J-C. (2005) Immediate gain is long-term loss: are there foresighted decision makers in Iowa gambling task? (Presentation at the third meeting of the society for neuroeconomics, Kiawah Island, September 2005.)

¹¹⁴ Bechara, A. & Martin, E. M. (2004). Impaired Decision Making Related to Working Memory Deficits in Individuals with Substance Addictions. *Neuropsychology*, 18(1), 152–162. This paper also showed that substance dependence subjects did worse at a memory task ('delayed non match to sample') than normal subjects with comparably poor IGT performance. In combination with other considerations this suggests the substance dependence is associated with compromised 'executive' cognition. See chapter 3 below.

The same has been found for patients on a methadone maintenance programme attempting to treat heroin addiction,¹¹⁵ and for long-term heavy users of marijuana.¹¹⁶

This suggests that, at least some of the time, the Iowa Gambling Task sorts relatively impulsive people into one pile and the others into another pile in the same way, roughly, as measures of discounting.¹¹⁷

A few other facts about the performance of different types of subject on the Iowa Task can be noted fairly quickly. In contrast to the results described above (in section 2-1) showing that cigarette smoking was clearly associated with steeper delay discounting, it seems as though worsened (or in any way changed) IGT performance is not associated with smoking.¹¹⁸ This result tentatively suggests that Iowa performance is to some extent independent of discounting. Nonetheless in a population of methadone-maintenance subjects, smoking was associated with further worsened performance at the IGT than was the case for non-smoking patients.¹¹⁹ This result tentatively suggests that Iowa Task performance, like discount rates, is cumulatively affected by multiple consumption problems, even if one of them is addiction to a substance, and the other disordered gambling.

A complication, though, is that the association between education and IGT performance seems to be roughly the opposite of the relationship in the case of delay discounting.

-
- ¹¹⁵ Mintzer, M. Z. and Stitzer, M. L. (2002). Cognitive impairment in methadone maintenance patients. *Drug and Alcohol Dependence*, 67, 41–51. For a result suggesting that the effect reported here is worse for subjects who in addition are smokers, see Rotheram-Fuller, E., Shoptaw, S., Bermanb, S. M. and London, E. D. (2004). Impaired performance in a test of decision-making by opiate-dependent tobacco smokers. *Drug and Alcohol Dependence*, 73, 79–86.
- ¹¹⁶ Whitlow, C. T., Liguori, A., Livengood, L. B., Hart, S. L., Mussat-Whitlow, B. J., Lamborna, C. M., Laurienti, P. J. and Porrino, L. J. (2004). Long-term heavy marijuana users make costly decisions on a gambling task. *Drug and Alcohol Dependence*, 76, 107–111.
- ¹¹⁷ It has also been found that anorexia nervosa sufferers do less well than controls. Anorexia is classified (by psychiatrists) as an ‘impulse disorder’. See Cavedini, P., Bassi, T., Ubbiali, A., Casolari, A., Giordani, S., Zorzi, C. and Bellodi, L. (2004). Neuropsychological investigation of decision-making in anorexia nervosa. *Psychiatry Research*, 127, 259–266.
- ¹¹⁸ Harmsen, H., Bischof, G., Brooks, A., Hohagen, F. and Rumpf, H.-J. (2005). The relationship between impaired decision-making, sensation seeking and readiness to change in cigarette smokers. *Addictive Behaviors*.
- ¹¹⁹ Rotheram-Fuller, E., Shoptaw, S., Bermanb, S. M. and London, E. D. (2004). Impaired performance in a test of decision-making by opiate-dependent tobacco smokers. *Drug and Alcohol Dependence*, 73, 79–86.

That is, having spent more years in education is associated with less steep discounting for delay,¹²⁰ but with worse performance at the IGT,¹²¹ even though in general IGT performance improves with age,¹²² even if some older subjects undergo a decline.¹²³ We also know that administration of testosterone leads to worsened performance on the IGT.¹²⁴

More significantly for our purposes, there is some work showing that disordered gamblers do poorly at the IGT compared to subjects from control groups. In one of these studies the disordered gamblers were found to be unimpaired at a different task intended to assess 'executive cognition' or the higher order cognitive functioning needed when the 'rules of the game' change and old responses have to be suppressed in favour of an alternative behaviour. (For example, remembering to call football 'soccer' while talking to a North American.) That the disordered gamblers were not impaired at executive cognition suggests that their poor performance on the Iowa Task was unrelated to executive cognition.¹²⁵ The other study found that the 'profile' of disordered gamblers (in a comparison of subjects with various sorts of decision making difficulties) was closer to that of alcoholics than sufferers of Tourette's syndrome.¹²⁶

-
- ¹²⁰ Jaroni, J. L., Wright, S. M., Lerman, C. and Epstein, L. H. (2004). Relationship between education and delay discounting in smokers. *Addictive Behaviors*, 29, 1171–1175. This study found that within a subject population of smokers, delay discounting steepness was inversely correlated with education level, that is that more educated subjects were less impulsive.
- ¹²¹ Evans, C. E. Y., Kemish, K. and Turnbull, O. H. (2004). Paradoxical effects of education on the Iowa Gambling Task. *Brain and Cognition*, 54, 240–244. This result may sound surprising, but since the IGT is supposed to measure a kind of 'emotional learning' the authors of the study hypothesised that more educated people would decide in more deliberate, and less emotional, ways, and hence do worse.
- ¹²² Hooper, C.J., Luciana, M., Conklin, H.M., & Yarger, R.S. (2004) Adolescents' Performance on the Iowa Gambling Task: Implications for the Development of Decision Making and Ventromedial Prefrontal Cortex. *Developmental Psychology*, 40(6), 1148–1158. There were three age groups in this study: 9–10 years, 11–13 years, 14–17 years.
- ¹²³ Denburg, N.L, Tranel, D., & Bechara, A. (2005) The ability to decide advantageously declines prematurely in some normal older persons. *Neuropsychologia*, 43, 1099–1106.
- ¹²⁴ Van Honk, J., Schutter, D.J.L.G., Hermans, E. J., Putman, P., Tuiten, A., Koppeschaar, H. (2004). Testosterone shifts the balance between sensitivity for punishment and reward in healthy young women. *Psychoneuroendocrinology*, 29, 937–943. In this study 12 healthy young women participated in a double blind, placebo controlled, crossover design. Subjects who had been administered testosterone made significantly more disadvantageous choices in later blocks of the IOWA Gambling Task.
- ¹²⁵ Cavedini, P., Riboldi, G., Keller, R., D'Annuncci, A. and Bellodi, L. (2002). Frontal Lobe Dysfunction in Pathological Gambling Patients. *Biological Psychiatry*, 51, 334–341.
- ¹²⁶ Goudriaan, A. E., Oosterlaan, J., De Beurs, E., and Van den Brink, W. (2004). Decision making in pathological gambling: A comparison between pathological gamblers, alcohol dependents, persons with Tourette syndrome, and normal controls. *Cognitive Brain Research*, 23, 137– 151.

It has also been found that problem gamblers perform worse than normal subjects at a task structurally similar to the IGT in key respects (the ‘game of dice task’ developed in Germany).¹²⁷ In this last study, though, unlike the first one mentioned in this paragraph, the frequency of bad choices made by disordered gamblers was found to be correlated with weaknesses in executive cognition.

As with the case of whether disordered gamblers differ from non-disordered gamblers in their delay discounting or their probability discounting (see section 2-2 above) the results just described need not be in conflict, even though they do initially seem to that way. There could, for instance, be different sorts of disordered gambler, some of whose problems are more related to executive cognition, and some of whose problems are not so related. In that case, the apparent conflict in results of the two studies could be because of neglected differences between their populations of disordered gamblers. Real differences between the IGT and the Game of Dice Task, alternatively, may mean that the apparent conflict disappears when experimental work on disordered gamblers using both tasks takes place, and we find that, and in what precise ways, they don’t measure the same thing.

Clearly enough, as already noted, some fraction of the responses of subjects when engaged in the IGT will be related to their discount rates. We don’t know how great a fraction partly because of the near-total dearth of experimental work assessing the same subjects in both ways, but we do know, from separate studies, that as well as performing poorly on the IGT, disordered gamblers are distinguished by their discounting. Some of these results have been mentioned already (in section 2-2 above). Here is a quick recap, along with some further results:

- It has been found that problem gamblers discount delayed rewards more steeply than non-gamblers.^{128,129,130}

¹²⁷ Brand, M., Kalbe, E., Labudda, K., Fujiwara, E., Kessler, J., & Markowitsch, H.J. (2005) Decision-making impairments in patients with pathological gambling, *Psychiatry Research*, 133, 91– 99.

¹²⁸ Dixon, M. R., Marley, J., and Jacobs, E. A. (2003). Delay Discounting By Pathological Gamblers. *Journal of Applied Behavior Analysis*, 36(4), 449–458. Subjects were considered pathological gamblers for this study if they had a SOGS score in excess of 4. The mean age of subjects in this study was 40.6 years.

¹²⁹ Alessi, S.M. & Petry, N.M. (2003) Pathological gambling severity is associated with impulsivity in a delay discounting procedure, *Behavioural Processes*, 64, 345–354. In this study steepness of delay discounting was found to be well predicted by SOGS scores. (Also, SOGS scores were a significantly better predictor than scores on the Eysenck impulsivity scale.) The mean age of subjects in this study was 44 years.

¹³⁰ Petry, N. M. (2001). Pathological Gamblers, With or Without Substance Abuse Disorders, Discount Delayed Rewards at High Rates. *Journal of Abnormal Psychology*, 110(3), 482-487.

- It has also been found that gamblers (not recruited as problem gamblers, but with SOGS scores in the problem gambling range) did *not* discount delayed rewards more steeply than non problem gambling control subjects, but were less sensitive to risk measured in a probability discounting task.¹³¹
- A measure of the attitude of risk, or perhaps the value of the future, and hence a measure of delay discounting, of a person is their willingness to engage in risky sexual activity. It has been found that substance abusers with gambling problems are more likely to engage in risky sexual behaviours than abusers without gambling problems.¹³²
- Cigarette smoking disordered gamblers have been found to have more severe gambling problems than non-smoking disordered gamblers, further confirming the association between smoking and disordered decision-making.¹³³
- It has also been found that disordered gambling substance abusers had dramatically higher discount rates (or steeper discount curves) than non-problem gambling substance abusers.¹³⁴ In connection with this, it is worth noting that the prevalence of disordered gambling is much higher among substance abusers than the population at large.¹³⁵

Although there is some possible conflict in these results, they are generally encouraging for the view that behavioural economics provides a suitable framework for making sense of disordered gambling. The very same research tools that distinguish drug abusers from others seem to distinguish disordered gamblers from subjects who are not disordered gamblers. This in turn suggests that disordered gambling is an example of a general kind of impulsivity we already know something about.

Further behavioural study of disordered gamblers is needed, though. As we've emphasised more than once, hardly any studies include gamblers of significantly varying severity of disordered gambling, and hardly any assess their subjects for both delay and probability discounting.

-
- ¹³¹ Holt, D. D., Green, L. and Myerson, J. (2003). Is discounting impulsive? Evidence from temporal and probability discounting in gambling and non-gambling college students. *Behavioural Processes*, 64, 355–367. The mean age of subjects in this study was 19.6 years. Bear in mind that discounting reduced probability less steeply shows risk insensitivity, the converse of the case with delay.
- ¹³² Petry, N. M. (2000). Gambling problems in substance abusers are associated with increased sexual risk behaviors. *Addiction*, 95(7), 1089–1100.
- ¹³³ Petry, N. M. and Oncken, C. (2002). Cigarette smoking is associated with increased severity of gambling problems in treatment-seeking gamblers. *Addiction*, 97, 745-753.
- ¹³⁴ Petry, N. M. and Casarella T. (1999). Excessive discounting of delayed rewards in substance abusers with gambling problems. *Drug and Alcohol Dependence*, 56, 25-32.
- ¹³⁵ Shaffer, H.J., Hall, M.N., & Vander Bilt, J. (1999) Estimating the prevalence of disordered gambling behavior in the United States and Canada: A research synthesis. *American Journal of Public Health*, 89, 1369-1376. See also: Petry, N.M. & Pietrzak, R.H. (2004) Comorbidity of substance use and gambling disorders. In H.R. Kranzler & J.A. Tinsley (Eds.), *Dual diagnosis and psychiatric treatment: Substance abuse and comorbid disorders* (2nd ed., pp. 437-459). New York: Marcel Dekker.

No study has ever included assessment on the IGT as well as at behavioural discounting procedures to begin shedding light on the relations between these instruments, both of which seem able to distinguish disordered gamblers from others.

Another crucial kind of research of which there is far too little concerns factors that temporarily influence discounting or choice behaviour. Almost all of the work above separates one group from another (heroin abusers from people who don't use heroin, or disordered gamblers from people whose gambling is unproblematic, etc.) measuring their performance at the experimental task, whether a discounting task or the Iowa, once off for the purposes of the study. We know, though, that factors of various sorts can induce temporary changes in how people make decisions. These factors include priming effects from seeing or thinking about something before making a choice, chemical effects arising from consuming various kinds of substance including alcohol, as well as factors like pain.

Starting with alcohol, it has been found that acute administration of alcohol leads to increased impulsivity in rats.¹³⁶ A study of real electronic gaming machine (EGM) players in a real venue found that, according to the EGM players, alcohol consumption was a major contributor to diminished control over whether to continue playing or not.¹³⁷ Using the questionnaire based instruments of personality psychology to measure impulsivity alcoholics, as well as the children of alcoholics, are more impulsive than the general population.¹³⁸

A different study in an experimental setting used real gaming machines, and measured the extent to which subjects persisted in the face of losses. Half of the subjects drank an alcoholic beverage before participating (equivalent to 30 grams of alcohol) while half drank a non-alcoholic placebo beverage. It was found that subjects who had consumed alcohol were twice as persistent against losses as members of the placebo group. Half of the subjects in the alcohol group lost their entire original cash stake, compared to only 15% of the placebo group subjects.¹³⁹

One question raised by this striking result is whether alcohol consumption leads to modification in discounting. In one very careful study all subjects were first assessed for delay and probability discounting, and then administered a beverage which in half of the cases was alcoholic while in other half it was a placebo, before being assessed again for delay and probability discounting. This study found that alcohol consumption had no effect at all on discount rates in the case of either delay or probability.¹⁴⁰

-
- ¹³⁶ Poulos, C.X., Parker, J.L., & Le, D.A. (1998) Increased impulsivity after injected alcohol predicts later alcohol consumption in rats: evidence for "loss-of-control drinking" and marked individual differences. *Behavioral Neuroscience* 112:1247–1257.
- ¹³⁷ Baron, E. & Dickerson, M. (1999). Alcohol Consumption and Self-Control of Gambling Behaviour. *Journal of Gambling Studies*, 15(1), 3-15.
- ¹³⁸ Graham, K. (1980). Theories of intoxicated aggression. *Canadian Journal of Behavior Science*, 12, 141–158.
- ¹³⁹ Kyngdon, A. and Dickerson, M. (1999). An experimental study of the effect of prior alcohol consumption on a simulated gambling activity. *Addiction* 94(5), 697-707.
- ¹⁴⁰ Richards, J.B., Zhang, L., Mitchell, S.H., & de Witt, H. (1999) Delay or probability discounting as a model of impulsive behavior: effect of alcohol, *Journal for the Experimental Analysis of Behavior*. 71, 121-143.

This result is, at least initially, very surprising, perhaps especially so given that alcoholics are known to be more impulsive in the specifically behavioural sense (see section 2-1 above).

It is possible, though, that the absence of a change in discounting in this case is explicable by reference to peculiarities of the procedures used to assess discounting. In them, subjects are repeatedly asked which of two options they prefer, but do not experience wins and losses during the procedure. Sometimes, at the end of the procedure, one or more of the choices are randomly selected and the subject paid whatever they said they preferred in the selected case. (This is done to encourage subjects to make the choices seriously, since any given choice could turn out to be 'real'.) In cases where the subject chooses a chancy option, they might well fail to earn anything, but their loss will take place only after the discounting procedure has been concluded, and hence have no effect on the procedure or its result.

More simply, if there are no losses in a standard behavioural discounting task, then one can hardly expect to see an effect of greater perseverance in the face of losses on the part of subjects who have consumed alcohol. This suggestion is supported by some of what is already known about the cognitive effects of alcohol consumption. Psychologists maintain that the effects of alcohol on social behaviours (at least) are more pronounced in situations containing both actual rewards and actual penalties,¹⁴¹ and especially where more substantial rewards and penalties were incurred sometimes.¹⁴² It is therefore clear that a comparative assessment of discounting behaviour under the influence of alcohol should be made in a situation with real consequences of choices, including choices leading to negative outcomes. Such an experiment would be easy enough to construct, either by determining the outcomes of at least some probabilistic choices in a discounting task right away, or by alternating choices in a discounting task with choices in some other task leading to losses and gains.

The findings that alcohol consumption leads to impaired decision making at gambling should be read in the light of additional work showing that gambling itself can lead to greater voluntary alcohol consumption (compared to a control activity of watching an action movie while alcohol was freely available), and that gamblers who were also drinking were more likely to undergo negative mood changes, even when gambling losses were factored in, than people who gambled but drank non-alcoholic beverages. That is, not only do people who drink and gamble make bad decisions, they are made less happy at the time they are gambling by the fact that they have been drinking.¹⁴³

¹⁴¹ For a review of over 30 laboratory studies see: Steele, C. M., & Southwick, L. (1985). Alcohol and social behavior I: The psychology of drunken excess. *Journal of Personality and Social Psychology*, 48, 18–34

¹⁴² Lau, M. A., Pihl, R. O., & Peterson, J. B. (1995). Provocation, acute alcohol intoxication, cognitive performance, and aggression. *Journal of Abnormal Psychology*, 104, 150–155.

¹⁴³ Stewart, S. H., McWilliams, L. A., Blackburn, J. R. and Klein, R. M. (2002). A laboratory-based investigation of relations among video lottery terminal (VLT) play, negative mood, and alcohol consumption in regular VLT players. *Addictive Behaviors*, 27, 819-835.

Leaving the case of alcohol aside, there are features of the external environment known to impact on decision. It has been known for some time that simply seeing a smiling face, even if so briefly that it is not consciously noticed or remembered, leads to greater reported approval of what is observed shortly afterwards.¹⁴⁴ Some very striking experiments in the early 1990s showed that the emotional content of sentences we consider can have appreciable effects on our behaviour. This work found that people who thought about sentences mentioning age walked more slowly for a period afterwards than those who didn't, and that thinking about sentences relating in different ways to assertiveness and politeness also tended to make the test subjects measurably more assertive or polite. (In these experiments the subjects believed themselves to be engaged in a language test with the sentences in question, and their rate of movement, or propensity for assertive or polite behaviour was measured in the corridor after they believed the experiment to have been concluded.)¹⁴⁵ A more recent study found that subliminal presentations of smiling faces before subjects were given the opportunity to pour themselves a drink led to pouring and consuming more of the offered beverage, and to increased willingness to pay for more of the beverage. Subliminal presentations of frowning faces had the opposite effect.¹⁴⁶

It is not known how, if at all, these manipulations and others like them impact on discounting, or on performance on tasks like the IGT. We do know, though, that some stimuli can make a difference to discounting. A few years ago two researchers hypothesised that attending to pictures of attractive members of the opposite sex would get male subjects into a 'mating mindset' and that one feature of this would be temporarily steeper discounting for delay. Their reasoning was that unless the present was valued disproportionately, opportunities would be sacrificed too often. They found just what they predicted, and also (although to a smaller extent) that looking at pictures of expensive cars had a similar effect on female subjects.¹⁴⁷

Some of the specific details of what is found in these sorts of studies may be surprising, just as the magnitude of the effect of some of the manipulations may be. The general point, though, is surely not in the least surprising: what people see around them makes a difference to what they do and how they feel about it. The ways that all sorts of places including public buildings, shopping centres, banks, hospitals and, of course, casinos are designed reflects this.

¹⁴⁴ Zajonc, R. 1980. 'Feeling and Thinking: Preferences need no inferences', *American Psychologist*, 35, 151-175. Zajonc, R. 1984. 'On the primacy of affect', *American Psychologist*, 39, 117-123.

¹⁴⁵ Bargh, J. 1990. 'Auto-motives: Preconscious determinants of social interaction', in T. Higgins and R. Sorrentino (eds.) *Handbook of motivation and cognition*, New York: Guilford. Bargh, J. 1992. 'Being unaware of the stimulus vs. unaware of its interpretation: Why subliminality per se does matter to social psychology', in R. Bornstein and T. Pittman (eds.) *Perception without Awareness*, New York: Guilford.

¹⁴⁶ Winkielman, P., Berridge, K. C., & Wilbarger, J. L. (2005) Unconscious Affective Reactions to Masked Happy Versus Angry Faces Influence Consumption Behavior and Judgments of Value. *Personality and Social Psychology Bulletin*, 31(1), 121-135.

¹⁴⁷ Wilson, M. & Daly, M. (2003) Do pretty women inspire men to discount the future? *Proceedings of the Royal Society of London (Supplementary – biology letters)*, 271, S177-179.

It is a matter for politics and policy, rather than science, to decide how much manipulation is acceptable advertising, and how much is problematically coercive. Science can help by trying to quantify the effects of different environmental factors, and the specific differences made to some kinds of subject (for example those at risk of being disordered gamblers) and other determinants in fluctuations in relative vulnerability to impulsivity. This knowledge should surely inform and influence both treatment and policy.

2.4. Conclusion

Abusers of substances and disordered gamblers both show up as relatively impulsive on behavioural and psychological measures of impulsivity, and also tend to perform poorly on some cognitive measures of decision. We need to know more than we do, which is presently rather little, about how the behavioural and cognitive approaches relate to one another.¹⁴⁸

The extent to which early impulsivity is a predictor of later disordered gambling is not known, although one suggestive study indicates that it is a predictor, and hence that individuals at risk of developing into disordered gamblers could be identified early. Longitudinal research is clearly required.

Although we know quite a lot about the effect of environmental factors and other consumption behaviour on some kinds of decision making, we know relatively little about what difference any of these factors (except alcohol consumption) make to gambling behaviour, and the risks of disordered gambling.

¹⁴⁸ We know of only one study in which Iowa Gambling Task performance and discounting performance were applied to the same subjects. A study a few years ago found that in a group of subjects (32 of them) presenting for treatment for cocaine dependence, that performance at the IGT and delay discounting were significantly correlated. Subjects were not assessed for probability discounting, and there was no control group in the study. Monterosso, J., Ehrman, R., Napier, K. L., O'Brien, C. P. and Childress, A. R. (2001). Three decision-making tasks in cocaine-dependent patients: do they measure the same construct? *Addiction*, 96, 1825–1837.

3. The Neuroeconomics and General Neuroscience of Pathological Gambling

EXECUTIVE SUMMARY

Two branches of neuroscience, neurochemistry and the newly emerged discipline of neuroeconomics, have jointly achieved major breakthroughs over the past few years in the understanding of addiction.

Neuroeconomics studies the behaviour of the brain as a calculator of the relative values of different possible and actual rewards that a person or other animal could pursue. Neuroeconomists have developed a rigorous and well-confirmed model of the computation of reward values, known as the 'reward learning / predictor valuation model'. Principal evidence for this model has been gathered by scanning subjects' brains under functional magnetic resonance imaging (fMRI) while they perform behavioural tasks under experimental control. This method has allowed the model to be tested in application to specific brain systems.

The most important of these is the so-called reward system, implemented by the joint action of an old part of the brain, the limbic system, and a part that has expanded with human evolution, the prefrontal cortex. This system integrates four main computational activities: (i) learning environmental cues that predict reward, (ii) learning comparative values of rewards, (iii) focusing attention on cues that predict rewards, and (iv) motivating the system to act on the basis of these cues. A finding of neuroscience that is surprising to common sense is that pleasure is only indirectly connected to reward, and is a relatively weak motivator of behaviour compared to other things, especially prospects for positive surprise.

Neurochemical and neurodynamical study of the reward system has broadly revealed how it works. Key to its operation are relationships among three main brain chemicals, or *neurotransmitters*. These are dopamine, serotonin, and GABA. Each is favoured by dedicated receptors in different parts of the system, and the arrangement of these receptors creates natural pathways by which the four computational processes above are carried out and integrated.

Drug addiction has been identified as, at the neurochemical and neurodynamical level, a pathology of the reward system. Different drugs exploit different pathways, with that exploited by stimulant drugs (cocaine, amphetamines) being the simplest, and those exploited by alcohol and opiates being more complicated. But all share general properties. They create flooding of dopamine into a particular part of the reward system. This causes the system to focus attention on the environmental cues that predict this flood, which are those associated with self-administration of the drug. The system then learns to orient toward these cues and is motivated to act on them. At the same time, glutamate levels are increased in prefrontal cortex and modify its neural connections. The modifications in question interfere with normal prefrontal inhibition of older brain systems and thereby cause behavior to become more impulsive. This facilitates drug-taking, and so addiction amplifies and locks itself in. The reward system becomes less receptive to alternative sources of reward. As addiction solidifies, the system learns to respond not to the drug itself, effects of which are readily predictable, but to the cues that predict it.

The neuroeconomic model explains why this all happens. Because the reward system is a *learning* system, it responds mainly to *surprising* events. This aspect of the system is responsible for the familiar pattern whereby an addict becomes *sensitized* to her drug, consuming ever more of it and with increasing concentration on it to the exclusion of other rewards.

A good deal of recent evidence strongly suggests that gambling, in some people, triggers reward system response that mirrors its response to the most direct form of chemical hijacking, that of stimulant drugs. This evidence is reviewed in depth in the chapter.

In light of what we now know about the reward system, and in light of the neuroeconomic model, it is less surprising that gambling should be a target of addiction. Gambling activities are *engineered* to produce rewarding surprise – that is, to produce exactly what stimulates the reward system. The gambler is buying a maximally convenient, direct manipulation of her mesolimbic reward system. If the result is neuroadaptation of prefrontal circuits to reduce inhibition of impulse, her brain is changed into the characteristic brain of the addict. She will experience cravings if she stops gambling, because other rewards will be less able to attract the attention of her reward system or motivate action. Thus she will be prone to relapse if she adopts a policy of trying to abandon gambling as its costs to her mount.

Fortunately, knowledge of the neurochemistry of addiction brings with it the promise of chemical interventions to counteract it. The chapter reviews very recent pilot studies – there being nothing more extensive as yet – that strongly suggest that AG can be effectively treated by drugs that disrupt addictive learning. These drugs do this by interfering with the neurotransmitter interactions that produce dopamine response amplification and sensitization. Of particular interest and promise are so-called ‘atypical antipsychotics’ that inhibit reuptake of dopamine in the heart of the reward system. The chapter concludes by indicating grounds for hope that within a few years there will be standard neuropharmacological therapies that allow addictive gambling to be effectively managed in the majority of cases, or perhaps even in all cases.

3.1. Neuroeconomics

People are not identical with their brains. By this we don’t mean to allude to any possible non-physical reality; that is an issue for philosophers and not relevant to our scientific ambit here. Rather, we just mean that insofar as people are identified with the totality of forces that influence their behavior and self-understanding, some of these forces reside outside their brains – especially in the patterns of cultural expectations and interpretations in which people participate. Let us illustrate what we are talking about by referencing it immediately to disordered gambling phenomena. In some cultural circumstances, a person might fit our definition of a problem gambler not because they’re compulsively drawn to gambling *per se*, but because they’re conflicted between accepting and rejecting some prevailing social norms, and gambling venues are convenient settings where these conflicts play out. Think, for example, of princes from conservative Arab monarchies who frequent European casinos. Some of these people might struggle to reduce their use of these facilities because of conflicts between social expectations on them and their own perceptions of themselves.

They would thereby fit the definition of a ‘problem gambler’. But we should not suppose that we could ever find a single kind of pattern in their brains that would be equivalent to the problem as experienced separately by each one of them. What they have in common with each other is a shared *social* problem that could express itself in different neural patternings from one individual to another.

Of course, the brain plays an essential *role* in the control of every behavior by every human being. Our point here is merely that we should not suppose in advance that phenomena we've grouped together under social considerations *necessarily* have anything interesting in common from a neuroscientific point of view. All human institutions give rise to some distinctive social problems that are *only* problems at the level where we categorize patterns by reference to social properties. Organized gambling institutions are no exception. Thus our broad category of 'disordered gambling' almost certainly includes some problems of this kind. Neuroscience should not be expected to be of much help in addressing them.

However, in this chapter we will explain why neuroscience *does* look like it will be the overwhelmingly important contributor to effective individual and public policy responses to an important sub-variety – one that may constitute the overwhelming majority of – *pathological* gambling. We stress that this is an empirical scientific discovery. Even if everyone might have expected that pathological gambling would have turned out to have a standard neuroanatomical or neuropsychological *component*, it was a perfectly reasonable hypothesis until very lately that it might also have had an irreducible social aspect that wasn't most usefully addressed by neuroscience alone. In fact, we think that as recently as two or three years ago this was still the *best* hypothesis. However, for reasons we will describe in this chapter, discoveries made and reported mainly since 2002 suggest very strongly that much or most pathological gambling is one expression (along with addiction to stimulant drugs such as cocaine and amphetamines) of a quite specific neural pathology.

Because the pathology is relatively specific, it lends itself to direct pharmacological response. We will argue that, thanks to these recent neuroscientific breakthroughs, there is real hope for the first time in human history that forms of addictive / compulsive behavior that have plagued us since at least the dawn of civilization could soon become as manageable as tooth decay or the control of large wild predators.¹⁴⁹ In the subsequent chapter we will comment on the implications of this for the organized gambling industry and the regulatory and health policies that govern it.

This is the chapter of our report that will involve by far the greatest number of scientific ideas and terms with which most readers will be unfamiliar. The chapter is mainly about the brain, including its parts and the chemicals that make it work. None of these parts or chemicals have everyday names, so it is simply impossible to talk about neuroscience in plain English. Yet we are claiming that neuroscience is about to revolutionize our understanding of and capacity for dealing with pathological gambling. It would therefore undermine the whole point of our work in this report to pass over the details because they can't be expressed non-technically. We are going to make an *argument* here about pathological gambling, and it is going to be based on *evidence*. Therefore, the reader must see what this evidence is so she can decide whether she thinks our argument is persuasive.

¹⁴⁹

We use these examples carefully. We are not saying here that pathological gambling can be made to go *completely* away. It is not in the nature of human problems to do that. In poor countries, and among very marginalized groups like untreated schizophrenics in rich countries, some people still *die* of infections caused by tooth decay. And of course crocodiles, lions, tigers, bears and sharks still kill, between them, several thousand people every year. The point of the comparison is instead this: tooth decay and large predators were once *major* human problems that could be coped with but not strictly *managed*. Now there are means by which, for every person, their probability of facing these problems *can* be reduced at easily bearable cost to nearly zero, though not all such probabilities for all people are *actually* so reduced, for various reasons. This is what we predict will become the situation, thanks to neuropharmacology, with pathological gambling (or, at least, with the incidence of pathological gambling that is *addictive* gambling; see below).

Faced with this challenge, of presenting evidence to non-technical readers about things and processes for which there are no words in non-technical language, we have adopted the following devices. First, at each stage where some technical terms are about to put in their initial appearance, we will introduce them explicitly. Terms being introduced in this way will be put in bold type. Second, in the parts of the chapter where we're presenting technical evidence, we'll follow each paragraph that states that evidence with a **non-technical synopsis** that states its main point in everyday English. We want to make clear that the reader trying to make up her mind about whether the evidence supports the conclusions we draw from it will not be able to do this just by reading these synopses. To reiterate: the evidence can't be put in everyday English because there are no everyday words for the objects of neuroscience. The synopses are for the reader who just wants to know *what* we are claiming and not *why* we are claiming what we do. Where readers with a main interest in policy and treatment are concerned, this might be exactly what they're looking for, and the synopses are for them. To readers who want to follow the argument and get a sense of why neuroscience portends such big changes in the shape of the pathological gambling problem, we suggest that they struggle with the unfamiliar words. For that is all that is truly technical here. We have tried to explain all deep *complexities* in terms of more straightforward ideas, and there are many subtleties we've left out as inappropriate in a document that isn't intended as a scientific paper. What will make the argument outside the synopses difficult reading are just new *words*, and we're going to introduce all of them. But once they start coming in thick batches, which can't be helped, then most readers aiming to follow the details will find they need to go slowly.

Straight off the bat we need to introduce vocabulary for the various branches of the sciences that are relevant to the brain mechanisms and processes underlying pathological gambling. **Neuroscience** refers indiscriminately to all of them together. **Neuroanatomy** studies the physical components of the brain, their placement relative to one another, and the mechanics of their interactions. (It is the part of neuroscience most analogous to auto mechanics.) When the reader looks at Figure 1 in section 3.2, the areas she sees labeled are neuroanatomical parts. Next: **neurochemistry** and **neurodynamics** study, respectively, the **molecular** and the **electrical / computational** bases by which these parts of the brain handle and transform information and pass it from one part to another. This refers to the fact that the brain does its work in two main physical ways. The first is by sending chemical signals that communicate about *overall conditions* – e.g., “Things are good here”, “Things are getting worse here”, “I am excited” – and *generic responses* – e.g. “I want to fight”, “I want to get away from this”, “I want you [another brain area] to stop talking to me”. The second is by sending electrical signals that communicate about specific states of affairs – e.g., “That is a red object”, “That is Mom”, “The capital of Zambia is Lusaka”.¹⁵⁰ In this chapter, the *details* we'll get into will mainly concern neurochemistry rather than neurodynamics. Finally, **neuropsychology** studies the ways in which neuroanatomy, neurochemistry and neurodynamics interactively influence perception, motor behavior, decision, thought, emotions, mood and consciousness. When you wonder about how your brain contributes to production of your thoughts and actions, identified in terms you can present to yourself in everyday language, you're mulling on neuropsychology.

¹⁵⁰ We warn the reader that putting things this way is a potentially misleading simplification. The meanings of signals within the brain don't literally have English translations. This is because English is a *public, communal* language that refers to things people have learned to coordinate on picking out *together*. What's 'discussed' inside the brain are things and states of affairs that *parts of the brain* have learned to coordinate on picking out together. A person can't 'see inside her head' and register these things consciously; so she can't label them. Thus types of 'brain objects' don't really line up, except (sometimes) very roughly and approximately, with types of 'personal objects'.

The disciplinary perspective from which we have written this chapter is that of a newcomer to the above set of neurosciences, one which will strike many readers as a bit bizarre on first encounter: **neuroeconomics**. We think that this is the most fruitful organizing platform from which to provide an integrated neuroscientific account of (much or most) pathological gambling. However, neuroeconomics is a sufficiently novel construct to most people's ways of thinking that we can't introduce it, as we did the neuroscience branches above, with just a few sentences. Instead, several pages of explanation, including explanation of the perspective's relevance to pathological gambling, are in order.

The basic assumption behind neuroeconomics is that natural selection couldn't leave all **valuation** that an intelligent organism¹⁵¹ must perform up to social-psychological processes. By 'valuation' we refer to the relative ranking of possible or actual states of affairs in terms of 'better than' and 'worse than'.¹⁵² Without an evolved internal currency in the nervous system, a creature couldn't assess the relative value of events relevant to expected supplies of different things it wants. For example, suppose an organism encounters a mating cue and a food cue at the same time, using different sensory modalities (i.e., information processed through different sense organs). To which shall it respond? Or, if it can respond to both, how shall it trade off possible relative allocations of attention to them? An organism that could only compare them by first working out a speculative theory of its own economic situation would likely fare very badly in any real environment. If it is to be guided by the evolutionary learning of its species – which is a very powerful source of information because it's been going on for so long – then its biological brain, not just its learned social / cultural contingencies, will need to play a role in informing its judgments. An organism that failed to make any use in its economic decisions of what its species had learned in evolution would be a very badly adapted – indeed impossibly badly adapted – organism. This means that the brain must directly compute over units that can represent the value of mating or eating (or fleeing or fighting or sleeping etc.) at some more abstract level – roughly as a person can weigh her house and her breakfast and the value of her life to her spouse in one currency, namely Dollars or Rands.¹⁵³

¹⁵¹ Here 'intelligent organism' just means: any organism that guides any of its behavior by reference to dispositions that aren't directly controlled by interactions of its genetic structure and the contingencies of a fixed set of environmental conditions that can be specified at the level of the organism's sensory organs. Very roughly: unintelligent organisms can be thought of as simple robots for which genetics made most of the important decisions in advance. Intelligent organisms take more of their own control 'inboard'.

¹⁵² When states of affairs are merely *ranked* this is called 'ordinal valuation'. When they're ranked *and* scaled with respect to one another – e.g., 'This is worth three times what that is worth' – this is called 'cardinal valuation'. Most of the valuation done by the reward system, the part of the brain that malfunctions in much or most pathological gambling, is cardinal valuation.

¹⁵³ People sometimes say that 'noble' goods like lives and attachments can't be valued in cash currency. This is a comforting sentiment to some, but it certainly isn't true in practice. No one pays infinite amounts for life insurance, and so specific, finite amounts have to be worked out. And most people would risk their lives running across a busy freeway for *some* amount of money.

So, what is this internal currency? Just a few years ago, no one had a very clear idea. Now, thanks to functional Magnetic Resonance Imaging (fMRI)¹⁵⁴ and other new technologies¹⁵⁵ that let us non-invasively scan groups of **neurons** (brain cells) in operation while people behave and think, we know.¹⁵⁶ Experiments performed initially with monkeys, but so far confirmed in humans, overwhelmingly suggest that the internal currency consists in variations in firing rates in dedicated neurons in parts of the brain that collectively make up the so-called '**reward system**' (that is, the system for processing information about actual or potential valuations). An increase in spike rates in these neurons means that a stimulus seems to the brain to be 'better than predicted', a decrease means it's 'worse than predicted,' and no change means it's 'just as expected' (so, this signal says to the rest of the brain, don't pay more or less than the default level of attention to it). Just to reassure the reader that this is a properly worked out scientific hypothesis, and not just a vague qualitative association, we can state, on the basis of work including carefully controlled observations over simian and human subjects, a 'predictor-valuation' model due to P. Read Montague and Greg Berns.¹⁵⁷ (Mathematics is coming! Again: the reader will not need to try to follow the details for the sake of things we'll say later. But we think it is important that she see the *form* of this model, so that she understands what *kind of* theory it is: it's a set of exact principles, like Newton's laws of motion. Thus we can exactly relate neuroeconomics to the behavioral economics discussed in Chapter 1. We think that if we didn't show the reader the form of the model, she wouldn't be able to understand why this is a kind of *economics*.) Suppose first that $R(x,n)$ estimates the value of a reward distributed at various possible times $x, y, z, \dots, n$ in the future, scaled according to the uncertainty attending to the intervals between the estimation point and each time as follows:

$$R(x,n; D) = \int_{-\infty}^{+\infty} dy G(x-y, (x-n)D)r(y), \text{ where } G(z,b) = (2\pi b)^{-1/2} \exp\{-z^2/2b\} \text{ and } D \text{ is a constant.}$$

¹⁵⁴ See Brewer, T., & Brown, J. (2003), Principles of pleasure prediction: specifying the neural dynamics of human reward learning, *Neuron* 38: 150-152. fMRI uses magnetic field properties to distinguish different patterns of blood flow in brains. Since we know a lot about the relationships between blood-flow patterns and the ways in which neuronal demands for blood vary with activity, we can use the blood-flow patterns to make reliable inferences about neural firing rates. In fMRI, these inferences are performed directly by the technology to yield statistical 'pictures' (not, as some people suppose, literal visual images) of neuronal patterns. The fMRI pictures only give us resolution down to the level of *groups* of neurons. But now researchers are discovering forms of statistical inference that allow the scientist – not the machine – to get from images of groups of neurons to (under some circumstances) conclusions about single neurons.

¹⁵⁵ Positron Emission Tomography (PET) has been around for a bit longer and is still the most useful neuroimaging technology for some purposes. Another new approach complementary to fMRI, but limited by its greater degree of invasiveness, is called Transcranial Magnetic Stimulation (TMS).

¹⁵⁶ See Glimcher, P. (2003), *Decisions, Uncertainty and the Brain*. Cambridge, MA: MIT Press; Kable, J., & Glimcher, P. (forthcoming) Neural mechanisms for valuing immediate and delayed rewards in humans. *Society for Neuroscience Abstracts*.

¹⁵⁷ Montague, P.R., & Berns, G. (2002). Neural economics and the biological substrates of valuation. *Neuron* 36: 265-284.

Then the following equation describes the value $F(n)$ the brain attaches to getting a particular predictor signal at perceptual time n :

$$\begin{aligned} F(n) &= \int_n^{+\infty} dx e^{-q(x-n)} \int_{-\infty}^{+\infty} dy G(x - y_1(x - n)D) \rho(y) \\ &= \int_n^{+\infty} dx \{e^{-q(x-n)}\} \cdot \{R(x, n; D)\} \\ &= \int_n^{+\infty} dx \{\text{discount future time } x \text{ relative to perceptual time } n\} \cdot \{\text{diffused version of} \\ &\quad \text{reward estimate } \rho(x) \text{ for some } x \text{ and } n\}. \end{aligned}$$

The predictor-valuation model provides the molecular foundation for the whole-system-level ('molar'; see Chapter 1) model of the kind of learning in which the reward system engages. This is so-called 'Temporal Difference (TD) Learning'.¹⁵⁸ Once again, we will give the equation, though (again) the reader will not need to try to learn it for the sake of understanding anything that comes later in the report. TD learning is a rule – an *algorithm* – that describes the way in which the reward system estimates a **value function** V (that is, a principle for taking a stream of possible information as input and turning it into a stream of advance relative value estimates as output). As McClure, Daw and Montague (2003)¹⁵⁹ describe it, this function “relates the situation at a particular time s_t to the expected, time-discounted sum of rewards (idealized as numeric measures r of received utility) that can be earned into the infinite future.”

Now suppose that $t, t+1, t+2$ etc. represent times on some arbitrary measurement scale. Then the TD equation is

$$V^*(s_t) = E[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \gamma^3 r_{t+3} + \dots]$$

which we close¹⁶⁰ by writing

$$V^*(s_t) = E[r_t + \gamma V^*(s_{t+1})]$$

where γ is a discounting parameter, as discussed in previous chapters of this report, between 0 and 1.

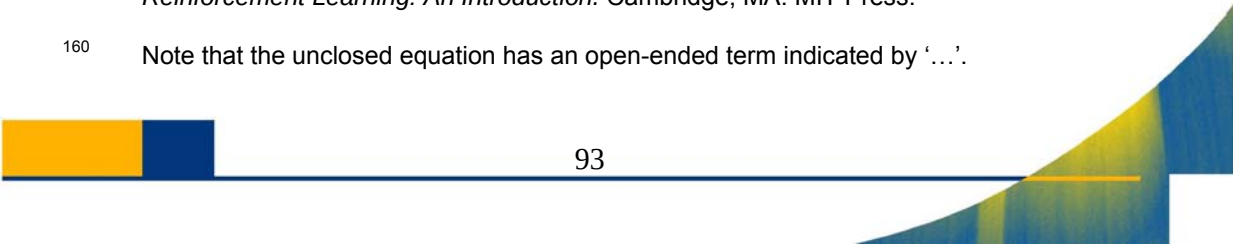
What this equation describes is the procedure by which the reward-system learning algorithm continuously inputs new information to keep refining its estimate of V to get a particular stream of actual temporal valuations V . From this we can define a measure δ of the extent to which the value estimates of two successive states and a reward experienced by the system are consistent with one another:

$$\delta(t) = r_t + \gamma V(s_{t+1}) - V(s_t).$$

δ is an error signal that pushes $V(s)$ towards better estimates as it gets more data. For example, if $V(s_{t+1})$ turns out to be better than expected, then $\delta(t)$ will be positive, thus indicating that $V(s_t)$ needs to be adjusted upward.

¹⁵⁸ McClure, S., Daw, N., & Montague, P.R. (2003). A computational substrate for incentive salience. *Trends in Neuroscience* 26: 423-428.

¹⁵⁹ As above, p. 425. For the general theory of TD learning, and its relationship to other forms of value learning used in economics and elsewhere, see Sutton, R., & Bartow, A. (1998), *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

¹⁶⁰ Note that the unclosed equation has an open-ended term indicated by '...'.


Learning that these molar and molecular functions describe the value-learning processes of the brain amounts to discovering the brain's economics, since we can use them to quantitatively model the opportunity costs the brain implicitly pays – in its own terms – for attending to one stimulus over another or issuing one **motor response**¹⁶¹ rather than another.¹⁶² This is why we refer to the study of the reward system using these parameters as 'neuroeconomics'. fMRI work has suggested that there are levels of discrimination in value-coding structure in the reward system from which the predictor-valuation model abstracts. (That is, the model averages over some of the details of reward representation at the molecular level.) Some neurons encode expected value (magnitude X probability) of rewards, while others encode a reward's subjective utility in context (i.e., its *opportunity cost* in terms of its cardinal ranking relative to other possibilities). It is likely that as neuroeconomics develops further over the coming years, it will call forth application of the full set of technical resources of mathematical microeconomics in understanding the principles by which the brain relates perception to the allocation of attention and preparation for action.

We recommend neuroeconomics as the framework for integrating the neuroscience of (much or most) pathological gambling, and addiction-like behavioral phenomena in general, because these are pathologies of the reward system. As we discussed in Chapter 1, they are patterns wherein the system *locally* values rewards that cause the *global* welfare of the person the system regulates to drop over time, and to regret her earlier choices despite the fact that she could (at least from some past point) have foreseen her own regret. Under what circumstances does $V(s_i)$ keep pushing the valuation of another gamble, or another hit of cocaine, higher than the available alternatives, even when some other aspect of the person – either some other system in her brain, or her whole self in aggregate – would prefer that she choose one of the alternatives? Just a few years ago the answers to this question were speculative. Now we know a good deal about them – though very far indeed from everything about them – on the basis of neuroscientific evidence. They are the subject of this chapter.

Before we go into the brain, however, let us pause to reiterate a few conceptual lessons driven home by the discussion in the previous chapters. If all people naturally observe the matching law (see chapter 1 section 2) and meliorate rather than globally maximize, then we don't need to resort to a specific reward system pathology to explain mere *problem* gambling. All we need for that is the fact that the person enjoys the gambling experience *at all* (some people don't), and that she receives that enjoyment in the short run. Then hyperbolic discounting alone will cause her to tend to regret the amounts she gambles except to the extent that she has learned to bundle her future choices with present ones, in the way described by Ainslie that we discussed in Chapter 1 sections 6 and 7. With respect to gambling (as with alcohol, marijuana and many potentially addictive targets) we know that most people successfully learn to bundle – we know it simply because most people manage to gamble non-pathologically and drink moderately. (Most also *occasionally* bundle less well than they wish they had awhile later, and many deliberately choose not to bundle from time to time, for reasons discussed in Chapter 1.)

¹⁶¹ A motor response is a brain signal ordering some part of the body to do something. (Bodies don't always obey, either because the body is damaged or because some other part of the brain simultaneously has conflicting ideas.)

¹⁶² For surveys see McCabe, K., (2003), Neuroeconomics, in Nadel, L., et al (eds.), *The Encyclopedia of Cognitive Science*, Nature Publishing Group; & Schultz, W., (2004), Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioral ecology, *Current Opinion in Neuroscience* 14: 139-147.

From the policy perspective, we address mere problem gambling by deciding to what extent we should institute measures that prevent gambling facilities from making it more difficult for customers to bundle, or that force them to help customers bundle a bit better. We will discuss this in Chapter 4.

What is shown by widespread participation in controlled gambling and drinking is that, at least with respect to these activities and some others, people can marshal resources stronger in effect than the brain processes we are about to review which dispose them toward addiction. Knowing this leaves many open questions. Are there opponent brain processes that, in most people, are sufficient by themselves to counteract addictive dispositions? Or is social reinforcement of moderation required? *Must* people make reference to socially constructed behavioral book-keeping strategies to maintain their personal (bundling) rules? (i.e., if Robinson Crusoe had a distillery on his island, and a limitless supply of fermentable material, could he avoid alcoholism?) Certainly, if social scaffolding is necessary to maintain personal rules, then there must be some brain processes that the scaffolding recruits and operates through; but this does not show that those processes could be adequate for the job on their own. It is relevant here that scientists in laboratories can bring *every* rat to addiction to any substance that is a target for addiction in *any* people; rat brains, at least, do not have adequate natural safeguards against their own addictive mechanisms. So perhaps our brains are like that too, and require social support to keep in order.

From an evolutionary point of view it is not surprising that brains on their own are vulnerable to runaway consumption of some rewards. Until our own ancestors developed agriculture, no animal could exercise enough control over its own supply schedule to need to control its addictive dispositions itself. When a group of elephants comes across fermented fruit, each one eats as much of it as he or she can and, if there's enough to go around, they all get intoxicated. But however much they would enjoy repeating the experience the next day or the day after, they do not have this option because they cannot cause fermentation to happen, or collect and maintain a reliable supply of their drug. Thus their ecological circumstances are a barrier against addiction, and there was no reason why natural selection would have wasted resources equipping them with an inboard one.

What about people? Since at least the dawn of agriculture, we have been able to control our supply schedules so as to allow for addiction to substances. By all evidence substance addiction has been a problem for some people since at least that far back.¹⁶³ Has natural selection responded by evolving a homeostatic (i.e., self-correcting, like a home air temperature system) neural mechanism that, in most people, blocks consumption of addictive substances after a certain threshold? Or did social learning gradually fill our cultural environments with normative reference points, customs and rituals that allow most people to avoid being enslaved by their own brain mechanisms? Were our early agricultural ancestors more threatened by addiction, before the relevant social scaffolding and/or homeostatic mechanisms evolved, than modern people? These are all questions for physical, social and economic anthropology. At present, we know some things that are relevant to addressing them – for example, that prefrontal cortex, which was the most recent part of the human brain to evolve, is responsible for controlling impulsive urges – but we don't know the answers. In particular, we don't know whether a genuinely asocial person who nevertheless (miraculously, all by herself) controlled her supply schedule could avoid addiction.

¹⁶³ See Rudgley, R. (1999), *The Lost Civilizations of the Stone Age*, New York: Free Press.

What about gambling? Pathological gambling, unlike substance addiction, does not require control over a scarce resource. So there is no obvious reason why pathological gambling, if some of it is addictive in the way that some substances are, had to wait for the development of agriculture. And, certainly, some contemporary hunter-gatherers gamble a great deal (by comparison with their devotion of time and energy to alternative pursuits, such as procuring food).¹⁶⁴ There is no evidence that non-human animals gamble, but this doesn't tell us as much as one might suppose. Wild animals with a taste for gambling would be behaviorally reckless animals, and such animals would generally perish quickly. Natural selection might therefore, just as in the case of dispositions to substance addiction, encounter no pressure to build neural safeguards against hijacking of animal reward systems by habitual gambling, since, again, ecological contingencies would ensure against such habits becoming common. Perhaps similar evolutionary logic applied to early hominids. To find out whether any animals are susceptible to addictive pathological gambling we would have to rely not on observation but on carefully contrived experiments, in which we deliberately put animals into non-natural environments where their possible addictive tendencies could flourish. To our knowledge, no experiments of the relevant type have been attempted.

Finally, in trying to come to grips with the relationship between social and neural factors for addiction, one might ask: are there some specific addictive targets for which socially reinforced personal rules are *generally* not sufficient? Nicotine is an obvious candidate here; comparatively few people smoke non-addictively. However, this is a far-from-decisive case for the existence of neural addicts that are independent of social learning. First, until fairly recently people who liked smoking had little reason known to them for bothering to resist addiction; there might well now be growing numbers of non-addicted (e.g. 'weekend' smokers) whose prevalence has not been carefully studied.¹⁶⁵ Second, even if nicotine is unusually addictive, this may be not because it is an unusually effective neurochemical hijacker of the reward system. Rather, it might simply be that the marginal cost of each individual cigarette smoked is so low, and is known by smokers to be so low, that people must develop uncharacteristically sophisticated personal rules before they can learn to control smoking, by which point addiction will usually have taken effect.

We have now reviewed salient points in our current scientific ignorance about the relationship between problems due just to hyperbolic discounting, against which personal rules alone can be effective, and so-called addictions. This general ignorance made it possible to doubt, until very recently, whether any pathological gambling, in the sense of the DSM-IV operationalization, is addictive in a neurochemical sense. That is, it was possible to suppose that pathological gamblers might simply be people who are, for reasons of social or cultural learning, especially incompetent at constructing personal rules. Since we have known for some time that some substances make people addicted by neurochemical mechanisms, this would have been one basis on which it could have been reasonable to deny that there is really such a phenomenon as *addiction* to gambling. If this is how the facts looked like coming out, then this survey could have stopped with the discussion of the behavioral economics of disordered gambling, and not required a long chapter on neuroscience.

¹⁶⁴ See Gray, P. (2004). Evolutionary and cross-cultural perspectives on gambling. *Journal of Gambling Studies* 20: 347-371.

¹⁶⁵ Such a study would be politically unpopular – and unpopular with granting agencies and journal editors – while public health authorities still have so much work to do in reducing smoking prevalence in general.

As we will now show, however, this is *not* how the facts have come out. In some people, gambling hijacks the reward system in the same way that stimulant drugs do. Neuroscientists have isolated many of the main details of the mechanisms and processes by which this happens. Thanks to their work, neuropharmacologists are in the course of developing chemical therapies that show great promise for reversing such hijacking. And there are grounds for optimism that they will soon be able to predict who is at risk for addictive gambling on the basis of a genetic screen. While non-addictive disordered gambling is probably something at which policy should aim only for some socially variable level of control (because that's the best that is possible except at outrageous costs), addictive gambling might well, in consequence of the recent breakthroughs in neuroscience, be something we can truly manage and aim to *minimize*. For reasons we will discuss in the final chapter, we think this has significant implications for industry participants and regulators.

One last point on which we must touch before narrowing our focus in this chapter to addictive gambling (AG) is: how much of pathological gambling is AG? The answer is that we do not know. The possibility that we might soon be able to diagnostically screen for AG has only opened up over the past couple of years, so it is still too soon to do the kinds of scientific surveys we'd need to do to estimate the prevalence of true AG among pathological gamblers as classified by DSM-IV. However, there are grounds for thinking that the proportion is very high – for thinking, indeed, that *all* pathological gamblers might be AGs. This is simply that, as we will document in section 3.5, empirical studies have found strong correlations between neurochemical properties of the reward system that have been identified with addiction and tendencies to pathological gambling. The pathological gamblers in these studies were of course identified using SOGS and similar screens. We know that these screens over-select for pathological gambling, as discussed in Chapter 1, section 1. Therefore, most samples probably contained people who, not being pathological gamblers, were surely not addicted gamblers (if there are any, which was the open question in these studies). But in that case, the proportion of *non-addicted* pathological gamblers in the samples that would have been consistent with finding strong correlations between addiction and pathological gambling has a low upper bound, because of the interference with statistical significance, the 'noise', already coming from the screening errors. Here, the tendency of SOGS to make errors in one direction as a screen for research purposes actually *helps* us make a useful inference. Most pathological gamblers, we will argue, are probably addicted, and it is possible that all are.

Because we assume that AG is a subset of pathological gambling – i.e., that no mere problem gamblers are AGs – we can without confusion use 'PG' here to refer to pathological gambling or gamblers (depending on grammatical context). Whenever there is a need to mention the distinction between pathological and problem gambling, both phrases will be spelled out. 'DG', as before, will refer to disordered gambling or gamblers (depending on grammatical context), and NADG to non-addicted disordered gambling or gamblers.

So, now, into the brain.

3.2. The Reward System

The organization of the chapter is as follows. Because AG is a pathology of the reward system, we must first follow the recent breakthroughs in neuroeconomics that have brought about our new understanding of this system. This will occupy the present section. In Section 3.3 we will then focus on addictive phenomena in general, so that we can later go on to understand AG as a specific – and highly revealing – manifestation of a general kind of reward system malfunction.

Sections 3.4 to 3.6 will zero in on our primary topic, AG itself. This will include discussion of AG as the simplest conceptual template (by which we mean, *basic form* for purposes of organizing our thought about more complicated forms) for addiction in general. In section 3.5 we will review the experimental neuroscience of AG, and in section 3.6 we will survey implications for drug-based treatment (that is, **neuropharmacological** treatment, i.e., based on drugs that work on neurochemicals).

Our first task, then, is to introduce the basic neuroanatomy, neurochemistry and neurodynamics of the reward system. A key general point is that the system integrates aspects of both an old, interior part of the brain, the **limbic system**, that we share with vertebrates in general, and the **prefrontal cortex** (PFC) whose dramatic growth (along with that of the frontal cortex) in hominids was a salient feature of our evolution and a necessary factor in our unique form of adaptive intelligence. Whenever evolved structures must co-opt the functional outputs of other structures that evolved much earlier, we should have our eye out for characteristic pathology patterns that result from the fact that the older structure was not selected for its capacity to serve the newer one. Natural selection cannot yoke together different systems evolved under different pressures and with different available building blocks, and avoid what engineers call 'kludges' – that is, incipient vulnerabilities that are patched around when the system is running under optimal conditions, but that expose predictable bugs when circumstances degrade in characteristic ways. The widespread and relatively standardized pathologies of addiction may well be kludges stemming from the imperfect adaptation of the limbic system to its role as a service department for newer cortical systems.

The figures below show the brain regions that play leading roles in the neuropsychology of reward. In the old interior part of the brain we have the **striatum**, which is itself part of the **basal ganglia** and includes the **nucleus accumbens**. It interacts intimately with the 'emotional centre', the **limbic system** that is composed of the **hypothalamus**, **amygdala**, and **hippocampus**. Below them are the 'basecamp for reward', the **ventral tegmental area (VTA)** and, flush beside VTA, the **pars compacta of the substantia nigra (SNpc)**. Just exterior to these, and separately evolved, is the **anterior cingulate cortex**.¹⁶⁶ In front and outside we have the recently expanded **prefrontal cortex**, divided into **ventromedial** and **dorsolateral** regions. At the bottom of it is a dedicated site of cortical reward processing, the **orbitofrontal cortex**. Reward processing and response patterns in people are functions of the ways in which these structures interact.

To help fix the idea of the various circuits and directions of influence we'll be describing, we begin with a series of images of the brain with areas of interest highlighted. The brain is structured, layered and folded in all directions, so no single image can do the job of showing what we need here.

In the first figure below the exterior of the brain is shown from the side, with the motor cortex, the dorsolateral prefrontal cortex (PFC) and the orbitofrontal cortex labeled. The ventromedial PFC borders the dorsolateral PFC, but lies on the insides of the two hemispheres of the brain, and so would not be visible from this orientation. The areas highlighted in this figure occur on both sides of the brain.

¹⁶⁶ Note the implication that our reward system in fact integrates not just two but three separately evolved structures. This doesn't just *add* potential kludges; it *multiplies* their extent and probability, since for every new unit we add to an integrated system, there's a new potential kludge for *each* interface with an existing system.

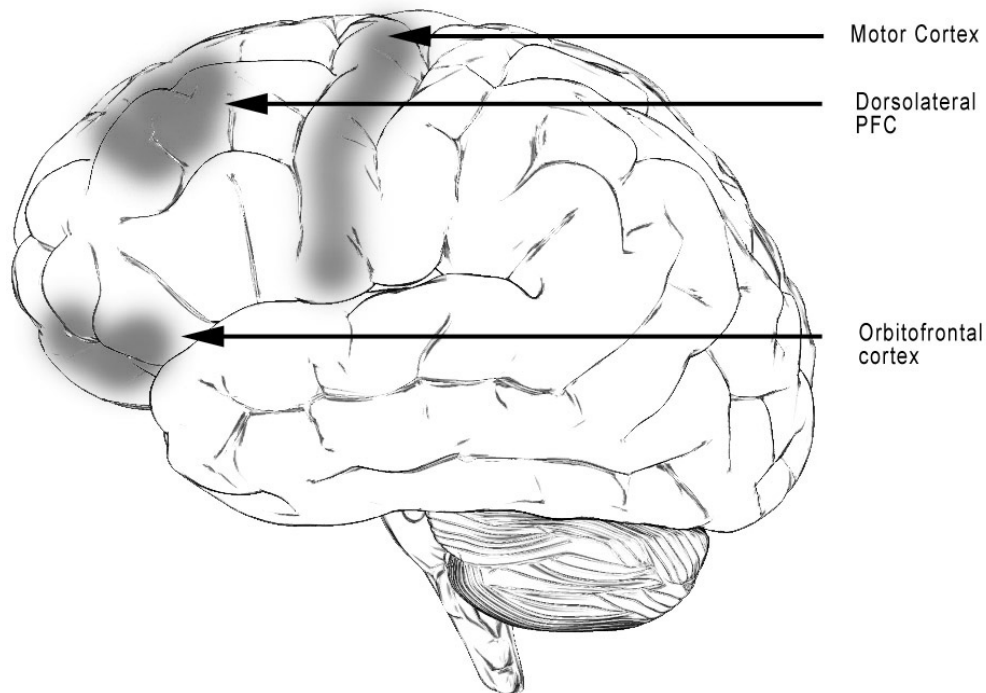


Figure 1

The main parts of the set of reward specialized brain systems we will be discussing are deeper within the brain. The second figure below is also a side view, but in it the one hemisphere has been removed, so that the middle of the brain ('facing' in the same direction as the image above, so that were the eyes included in either image they would be looking to the left) is being viewed. The substantia nigra is located close to the ventral tegmental area appearing on this figure:¹⁶⁷

¹⁶⁷

Figure adapted from Laviolette, S. R. & van der Kooy, D. (2004) The neurobiology of nicotine addiction: bridging the gap from molecules to behaviour, *Nature Reviews Neuroscience*, 5, 55-65.

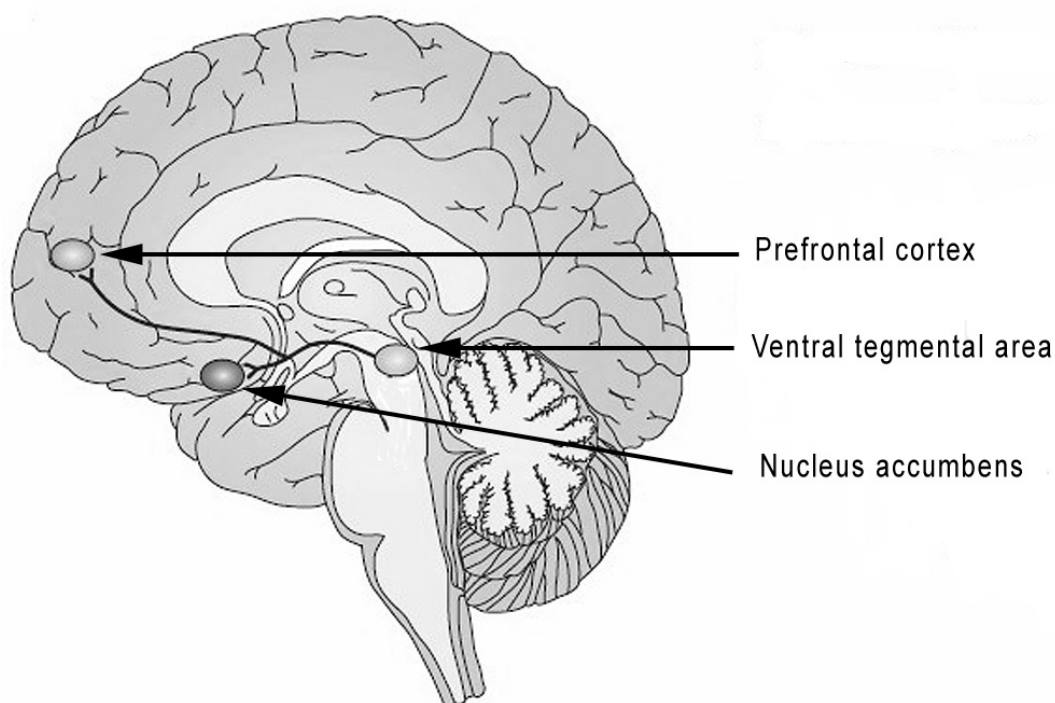


Figure 2

The third figure shows the brain in a more roughly three-dimensional view, as though it was facing in the direction of the thick black arrow. Note that the hypothalamus borders the thalamus, and does not appear on this figure because it is obscured by the dorsal striatum.¹⁶⁸

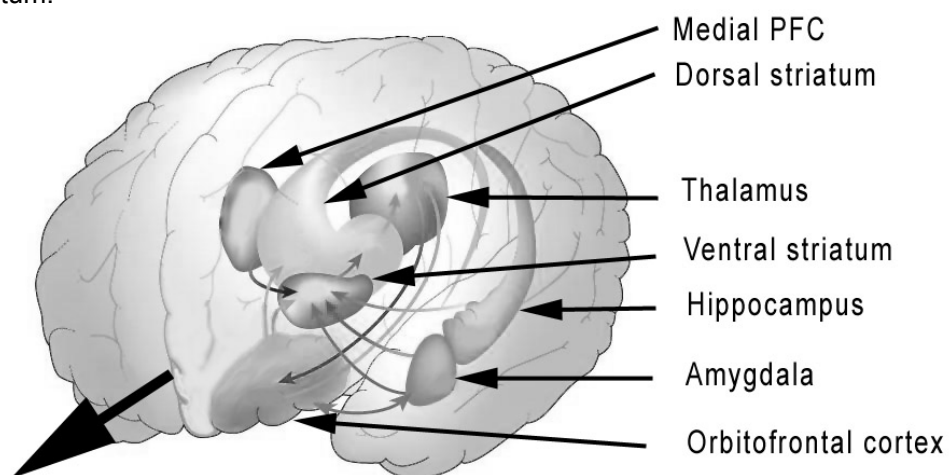


Figure 3

¹⁶⁸

Figure adapted from Everitt, B. & Robbins, T.W. (2005) Neural systems of reinforcement for drug addiction: from actions to habit to compulsion, *Nature Neuroscience*, 8(11) 1481-1489.

Our final figure is schematic, sacrificing attempt at fidelity to anatomical proportions in order to represent a system of connections. Some of the terms appearing in it ('dopamine', 'GABA') are brain chemicals (neurotransmitters) discussed later in this chapter.¹⁶⁹

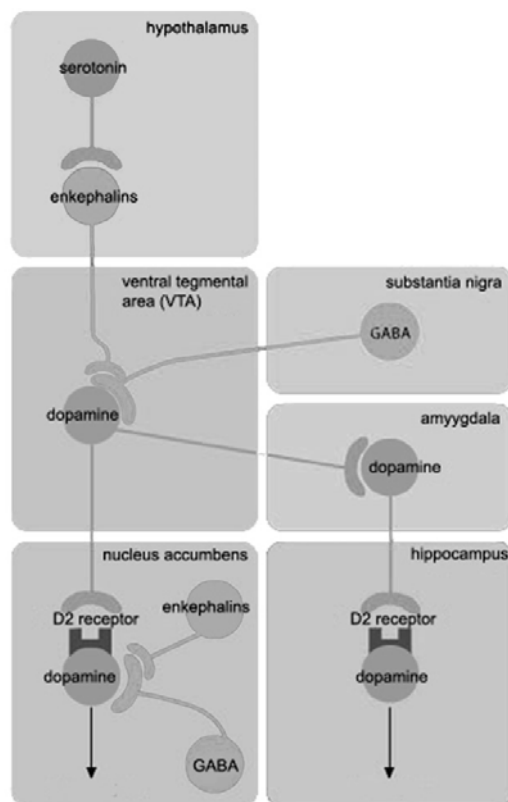


Figure 4

It has been recognized for many years, mainly on the basis of work with animals, that the VTA and the SNpc are the two structures that send reward signals out through the rest of the brain. Their role was revealed in a tradition of experiments whereby rats who were supplied with levers they could press to directly stimulate these areas through implants would do so repeatedly and relentlessly, ignoring less direct reinforcers like food and water to the point of death.¹⁷⁰ In this pattern researchers immediately recognized the prototype for addiction. The discovery of its underlying neurochemistry and neurodynamics has constituted the main part of the history of reward-system research. Relevant facts based on animal work have been under assembly for decades; but development of the integrated neuroeconomic model had to await fMRI investigation since 2000.

We distinguish VTA and SNpc from one another mainly by reference to the different **pathways** through which they feed the reward signals forward. These signals consist in release of a **neurotransmitter** (see next paragraph) called **dopamine**.

¹⁶⁹ Figure adapted from "The Brain from Top to Bottom": http://www.thebrain.mcgill.ca/flash/i/i_03/i_03_cl/i_03_cl_que/i_03_cl_que.html

¹⁷⁰ Niesink, R., Jaspers, R., Kornet, L. & van Ree, J. (eds.) (1999). *Drugs of Abuse and Addiction: Neurobehavioral Toxicology*, chs. 5 and 6. Boca Ration: CRC Press.

It is through variance in dopamine-releasing rates that VTA and SNpc implement the predictor-valuation model described in the previous section.¹⁷¹ Above-baseline dopamine release is interpreted by the rest of the brain as meaning that a reward has been received that was better than expected, disappointment relative to expectations is indicated by below-baseline dopamine release, and baseline dopamine release is the default state meaning that reward levels are as predicted. It will be crucial to our understanding of addiction and PG that the brain in general is a system that responds to *changes*; it is not like a monitor of a film that explicitly registers everything going on.¹⁷² It *attends* to surprisingly good and surprisingly poor rewards and tends to ignore conditions that don't change the status quo. The large human brain is primarily an organ adapted for *learning* – to which status quo conditions make no contribution. Therefore, brains do not waste resources on them. The reward system is no exception to this principle. Though common sense is not used to thinking of reward-processing as a form of learning, that is what it is and that is how it is neurally implemented.¹⁷³

We will have a great deal to say about **dopamine** throughout this chapter, so before we proceed further we should offer an explanation of **neurotransmission** and the **neurotransmitters** that perform it. One of the features of the brain that most contributes to its complexity is that it uses a two-dimensional information transmission system. One dimension is electrical. Brain cells are wired together in circuits through which varying electrical pulses are passed along connecting cables called **dendrites**. Different brain cells participate in different neural pathways partly by means of their various specific dendritic connection patterns. However, neurons don't transmit the signaling pulses from one to another by implementing an electrical switchboard. Instead, they influence one another's probability of receiving signals by emitting chemicals at **synapses**. How likely it is that a given signal will cross the space between the synapse of one neuron and the synapse of another neuron is a function of the level of that synapse's chemical of interest in the fluid of the brain at the site in question. Neurons are sorted into types on the basis of which neurochemicals their synapses release and respond to. These neurochemicals are called '**neurotransmitters**', and so neurons are typed after the neurotransmitters in which they deal. For example, thanks to the recent popularity of anti-depression agents like Prozac that modulate **serotonin** levels in different parts of the brain, serotonin has become the best-known neurotransmitter among the general public. We then group all the parts of the brain that communicate with one another by releasing and taking up serotonin as together comprising the **serotonergic system**.

¹⁷¹ For an early specification of the model in application to the dopaminergic system, see Schultz, W., Dayan, P. & Montague, P.R. (1997), A neural substrate of prediction and reward, *Science* 275: 1593-1599.

¹⁷² Popular confusion on this point is the main reason that consciousness is philosophically perplexing to most people. Since many people reading about neuroscience but new to it immediately wonder about consciousness in connection with it, a brief note is perhaps helpful here. We are incapable of noticing, from the inside, the difference between our direct pickup of change signals and our indirect, but equally automatic (when probed), interpretative judgments of 'no change' states. There is no reason natural selection should have built us to be able to notice this difference, so it didn't. But this entails that we cannot try to understand consciousness by introspection, by reflecting on what it feels like to be ourselves. Understanding consciousness requires doing neuroscience – contrary to what many philosophers believe and assert. This point is made in an accessible and vigorous way by Dennett, D. (1991), *Consciousness Explained*, Boston: Little Brown, and by Churchland, P. (1996), *The Engine of Reason, The Seat of the Soul*, Cambridge MA: MIT Press.

¹⁷³ This came as no surprise to psychologists trained in the behaviorist tradition.

Dopamine is the neurotransmitter we will have most to say about in this survey. Following the previous example, the pathways through the brain created by linked dopamine releasers and receivers are referred to as the **dopaminergic system**. We apologize to the reader for the fact that this talk of ‘systems’ – for which we aren’t responsible – wasn’t designed to help lay people keep easy track of things. We will refer to three neurotransmitter systems, the dopaminergic, the serotonergic and the **noradrenergic**, in this report. The reader will get confused if she thinks of the reward system as a *separate* system from these neurotransmitter systems. The latter are distinguished from one another neurochemically. By contrast, the reward system is distinguished *neuropsychologically*, that is, by reference to its role in relating the whole organism’s perceptual uptake and its behavior. Consider, by analogy, the training system and the player-ownership system in rugby: neither system is a *part of* the other, but they are related, and the same players participate in both systems. So, similarly, the neurotransmitter systems play distinct roles *within* the reward system, while also playing other, complementary, roles in other neuropsychological systems such as the language-processing system. One and the same neuron may be part of the dopaminergic system from the neurochemical point of view, part of the limbic system from the neuroanatomical point of the view, and part of the reward system from the neuropsychological point of view.

The reward system is the set of processes in the brain that computes the outputs broadcast as dopamine pulses by VTA and SNpc. We will begin our tour of it with the hypothalamus. This directly monitors basic bodily states, such as blood sugar and temperature levels, that are direct proprioceptive (i.e., inwardly but unconsciously perceived) indicators of changes relevant to the overall system’s present and expected utility. Hypothalamic information serves as fundamental input to the reward system, and *connects* reward to representation of **hedonic pleasure**,¹⁷⁴ which (simplifying somewhat) is the business of a part of the anterior cingulate cortex. We have italicized ‘connects’ here to stress that although reward and pleasure are related, they are not the same thing. This may be the single most important point on which neuroscience calls for a revision in popular ‘common sense’ about addiction and compulsive behavior. When reward and pleasure are muddled together, as they are in everyday thought and talk we’re led to think that (e.g.) a heroin addict must pursue heroin ‘because she likes it’. Then her attempt to kick her habit seems to require us to say that she simultaneously *doesn’t* like it – and now we find that it’s hard to say anything coherent about the whole matter. Neuroscience forces us to separate not just two but three kinds of psychological state: (i) the extent to which a state is hedonically pleasurable; (ii) the extent to which a *change of state* is treated by the brain as a reward; and (iii) the extent to which a person *judges* a change of state as improving or reducing their immediate or longer-term welfare. It turns out from the study of the brain – i.e., this is not something one settles by appeal to philosophical opinion – that (ii) and (iii) are more intimately associated with one another than *either* is with (i). A landmark fMRI study suggesting this was performed and reported by Berns and colleagues in 2000.¹⁷⁵ They found that subjects’ subjective reports of pleasure discrimination between fruit juice and water were correlated only with activity in cortical areas associated with sensory processes, whereas responses to unpredictability were correlated with activity in nucleus accumbens (Nacc), thalamus and medial orbitofrontal cortex (OFC).

¹⁷⁴ This refers to nice sensations. The point of ‘hedonic’ is to distinguish this sort of pleasure from cognitive pleasure, such as being ‘pleased that’ South Africa has many languages and cultures.

¹⁷⁵ Berns, G., McClure, S., Pagnoni, G., & Montague, P.R. (2001). Predictability modulates human brain response to reward. *Journal of Neuroscience* 21: 2793-2798.

These areas have in common that they are involved in dopamine-regulated input to attention allocation and (through connection with motor neurons) to preparation for action. Thus pleasure response seemed to be occurring outside the direct reward-response loop.

A wave of subsequent experiments building on this classic investigation¹⁷⁶ have established as a basic principle of neuroeconomics that people (behaviorally and neurally) ‘want’ many things they don’t (hedonically) ‘like’.¹⁷⁷ Furthermore, they can readily learn not to want things that they like, if the latter become predictors for other things they don’t want or like. For example, people who successfully lose weight often do so by *learning* to find fatty foods that they formerly enjoyed disgusting to consider. Of course, many targets of addiction induce hedonic pleasure while they’re being consumed, and this is often crucial to their *initial* attractiveness. However, just as many smokers actually find the sensations of at least many of the cigarettes they smoke (e.g., the 30th of a long night) *mainly* unpleasant (while smoking them anyway), so we should not assume that a pathological gambler *must* experience a warm subjective thrill while she’s rolling the dice. Hedonic pleasure is so incidentally and unreliably connected to reward *and to the motivation of behavior* that we’ll have little further to say about it in this report.

Let us add just a bit more on this theme, since, as noted, it is probably the most important point on which scientific and everyday conceptions of addiction and compulsion diverge. There is an obvious reason why most people don’t clearly distinguish between pleasure and reward, even though they play such distinct roles in the brain. Consciously experienced pleasure is *one* (only somewhat reliable) proxy for, or predictor of, reward. Precisely *because* it is the consciously experienced signal of probable reward, it’s one we notice when we introspectively reflect on the motivations for our behavior. (We also notice some motivators programmed into us by socialization, such as injunctions to moral duty; these get coded in public language and we notice them by literally talking to ourselves as if we were other people, not by directly ‘looking inside’.) However, most of what goes on in our brains is not consciously accessible in itself, while nevertheless contributing to our *interpretations* of what we *think* we’re experiencing. Hedonic pleasure signals evolved in primitive ancestors of ours that had much less complicated reward systems than we do, and the pleasure system we’ve inherited from them probably responds unequivocally – that is, in a way not strongly mediated by reward signals – only to some very basic cues, such as air temperature, whether we are damp or dry, and sexual contact. The rest of what we loosely *interpret* as pleasure is probably built out of second-order associations between pleasure signals and reward signals (e.g., getting into bed predicts warmth and sex, so getting into bed comes to be experienced as directly pleasurable). As creatures whose main evolutionary adaptation has itself been adaptive intelligence, however, we have reward systems that were precisely selected so as *not* to give pleasure signals priority over our curiosity about changes in environmental reinforcers. This is why pleasure is not a very *reliable* proxy for reward in people. More to the point, as emphasized at the bottom of the preceding paragraph, it isn’t very important to behavioral motivation.

¹⁷⁶ A ‘classic’ already, from 2000 – and a paper that still had to speculate hopefully on a number of conjectures that are now among the fundamental starting facts of reward-system neuropsychology. Discoveries are occurring very quickly in this area.

¹⁷⁷ A very simple example is that many people learn to want – in the sense of behaviorally and neurally treating as rewarding – very hot peppers that are coded by anterior cingulate cortex as aversive.

People who suffer catastrophes to their anterior cingulates and report losing hedonic color from their subjective worlds still behave in broadly normal ways.¹⁷⁸ By contrast, lesions to the components of the reward system typically produce profound breakdowns, ranging from Parkinson's disease to total motivational disintegration. Rats whose reward systems were trained to respond positively to terrible-tasting food ate the food while making facial expressions characteristic (in rats) of unpleasant experiences.¹⁷⁹ The fact that we are evolved to pursue what stimulates our reward system more diligently than what makes us feel good will turn out to be a fundamental part of the neuroscientific explanation of addiction and AG.

Returning from our digression on the pleasure/reward distinction to our tour of reward-system neurochemistry and neurodynamics, the hypothalamus dynamically represents (i.e., keeps updating with changing events over time) the visceral state of the body to limbic and cortical ('cognitive') systems. Each of these systems uses this basic data in its own distinctive way, and neuroscience is still some way from having an integrated account of the whole network. We know, of course, that at the *molar* level we get the appearance of hyperbolic discounting of utility, often controlled by development and maintenance of personal rules, as described in Chapter 1. Various molecular hypotheses are consistent with this molar account.

In a 2004 paper McClure and colleagues¹⁸⁰ report on an fMRI experiment that compared subjects offered choices that included offers of immediate rewards with subjects choosing only amongst more delayed prizes. They observed increased activity in limbic-system (ventral striatum and medial OFC) neurons when choices included earlier rewards, and increased PFC activity whenever subjects actually chose later, larger rewards. On the basis of this, they hypothesize that molar hyperbolic discounting is an abstract description of the aggregated composition of two separate discounting mechanisms lodged in different brain systems. Specifically, they "hypothesize that short-run impatience is driven by the limbic system, which responds preferentially to immediate rewards and is less sensitive to the value of future rewards, whereas long-run patience is mediated by the lateral prefrontal cortex and associated structures, which are able to evaluate trade-offs between abstract rewards, including rewards in the more distant future". In earlier work, Laibson¹⁸¹ shows how to technically recover *some* of the behavioral data explained by hyperbolic discounting by means of curves that combine two exponential curves of different slopes. He calls this ' δ - β discounting', where ' β ' ('beta') denotes a steep-sloped function corresponding to discounting of immediate prospects and ' δ ' ('gamma') denotes a less steep-sloped function corresponding to discounting of more delayed prospects. Then the hypothesis proposed by McClure and colleagues is that lateral prefrontal areas (which they called the " δ system" after Laibson's model) are active whenever subjects make any decisions.

¹⁷⁸ Berridge, K., & Robinson, T. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Research Reviews* 28: 309-369; Schroeder, T. (2004). *Three Faces of Desire*, pp. 121-122. Oxford: Oxford University Press.

¹⁷⁹ Berridge, K., & Valenstein, E. (1991). What psychological process mediates feeding evoked by electrical stimulation of the lateral hypothalamus? *Behavioral Neuroscience* 105: 3-14.

¹⁸⁰ McClure, S., Laibson, D., Loewenstein, G., & Cohen, J. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science* 306: 503-507.

¹⁸¹ Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics* 62: 443-478.

The limbic (“ β system”) is active at a level comparable to that of the δ system (“(and with a trend toward greater β -system activity)”) when subjects choose earlier options over later ones.

Non-technical synopsis: *Some researchers have suggested that brain systems don’t hyperbolically discount. Instead, they argue, the outside of the brain discounts exponentially and the inside of the brain does so too, but more steeply – it cares less about the future. Overall human (and other animal) behavior just seems hyperbolic because at that level our measurements average over cases where the outside of the brain is in charge and cases where the inside is in charge.*

As Ainslie and Monterosso¹⁸² point out, this hypothesis outruns the data. The evidence from increasingly many studies indeed tells us that elevated limbic activity and decreased or impaired prefrontal activity are independently correlated with impulsivity. But the key word here is ‘independently’. (This could not emerge given McClure and co.’s experimental design; we will encounter the evidence in studies to be reviewed shortly.) We are not in a position to claim that we have *rival* centres of discounting that *compete* for control of molar choice. This would greatly complicate the existing model of the reward system, as Ainslie and Monterosso point out: “Although the increased activity of ... lateral prefrontal cortex ... in response to larger/later selections is an important finding, to accord these areas status as a separate decision-making mechanism would add a complicating factor that would have to be reintegrated with motivation”. Kable and Glimcher¹⁸³ have new fMRI data that we interpret as showing that, in fact, neurons in limbic and prefrontal areas implement similar (hyperbolically shaped) discount functions. If this is right, then bundling must, as Ainslie argues, involve incorporating future-indexed expected utility into immediately indexed utility. We could then say, if we like, that all DG episodes involve failure to so incorporate – that is, failure to bundle. Mere problem gambling involves (more or less often) occasional failure to bundle, for circumstantially contingent reasons that vary from occasion to occasion. PGs are people who *cannot* bundle where gambling is concerned. We should look directly to neurochemistry and/or neurodynamics to tell us what prevents them from doing so.

Non-technical synopsis: *Even the smallest parts of the brain, and not just the whole person or other animal, seem to discount hyperbolically. Thus all avoidance of addictive-style behavior, even by parts of the brain, must involve bundling, and so there must be a neuroscientific account of it. Therefore, if PGs are people who can’t bundle no matter how hard they try, we should look in their brains to explain why their neural bundling mechanisms aren’t working.*

Other hypotheses about specific contributions of different brain systems to overall behavioral valuation remain open. Glimcher and Rustichini¹⁸⁴ suggest that ventromedial prefrontal cortex attaches so-called ‘somatic markers’ – roughly, emotional biases – to outputs of the reward system that cause most people to fail to behave like expected-utility maximizers (see Chapter 1, section 3), even in the very short run, under certain kinds of situations.

¹⁸² Ainslie, G., & Monterosso, J. (2004). A marketplace in the brain? *Science* 306: 421-423.

¹⁸³ Kable, J., & Glimcher, P. (forthcoming) Neural mechanisms for valuing immediate and delayed rewards in humans. *Society for Neuroscience Abstracts*.

¹⁸⁴ Glimcher, P., & Rustichini, A. (2004). Neuroeconomics: the consilience of brain and decision. *Science* 306: 447-452.

They discuss evidence, for example, that most people are averse to losses, in the sense that they'll prefer an expected reward of a given magnitude to another of identical magnitude if the latter has a higher probability of inflicting a loss, and that they'll pay (in expected utility) to avoid ambiguous choices. Glimcher and Rustichini cite observations suggesting that people with damage to prefrontal brain areas behave *more* like (short-run) expected-utility maximizers in these conditions. This speculative possibility is among several that merit further research, and is relevant to predicting the behavior of NADGs under different sorts of circumstances. There is no evidence, however, that the explanation of addiction or obsessive-compulsive behavior lies in these complications. They may account for respects in which healthy gamblers don't behave like perfectly rational neoclassical economic agents, but they do not appear to be part of the basis of AG.

Non-technical synopsis: *There's evidence that emotional fear of loss might play a role in protecting people from NADG. But damage to emotional systems doesn't seem to be a direct part of the basis for addiction (or, therefore for AG).*

So let us focus more specifically on the reward system, *within* which, as we will see, the basis of AG seems to lie. A first key finding is that unpredictable stimuli produce stronger responses in VTA, Nacc and ventral striatum than do familiar ones.¹⁸⁵ A second finding is that the stronger Nacc response is a direct function of an elevated surge of dopamine to it from VTA. This response specifically signals *positive* surprises; dopamine neurons show only minor, probably indirect, response to aversive and neutral stimuli.¹⁸⁶ In particular, Nacc is unresponsive to such stimuli.¹⁸⁷ Thus, contrary to what common sense might have expected, reward and punishment are not coded along a single continuum to which an integrated brain system responds. (This will prove important to understanding why adverse consequences of addiction and AG are notoriously ineffective, at least by themselves, in reconditioning afflicted people.) Dopamine neurons do not discriminate with respect to the sensory modality through which stimuli are delivered; speaking anthropocentrically, they 'care' only about the extent to which a stimulus is relevant to the difference between the prediction and the delivery of its associated reward.¹⁸⁸ (This is important to explaining why, in addicts and AGs attempting recovery, *anything* that has been a characteristic part of the person's habitual environment for drug-taking or gambling is an equally strong potential trigger of relapse.) As learning proceeds, the system stops responding to administration of the reward itself and responds instead to its predictors.¹⁸⁹

¹⁸⁵ Berns, G., McClure, S., Pagnoni, G., & Montague, P.R. (2001). Predictability modulates human brain response to reward. *Journal of Neuroscience* 21: 2793-2798.

¹⁸⁶ Mirenowicz, J., & Schultz, W. (1996). Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature* 379: 449-451; Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron* 36: 241-263.

¹⁸⁷ Knutson, B., Adams, C., Fong, G., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *Journal of Neuroscience* 21: RC159.

¹⁸⁸ Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron* 36: 241-263.

¹⁸⁹ Schultz, as immediately above.

Furthermore, dopaminergic neurons show depressed response when rewards are smaller than what learning would have led the system to predict; thus the dopamine neurochemistry converges qualitatively with the predictor-valuation model that quantitatively describes the fMRI data.¹⁹⁰

Non-technical synopsis: *The limbic system gets surges of dopamine when it gets positively surprised. The part of the limbic system most responsible for addiction doesn't get less dopamine than normal when it's negatively surprised. So, in the brain, punishment isn't just the opposite of reward, and we shouldn't think we can cure addicts by punishing them. It doesn't matter to dopamine release how the body gets the surprising information, so any clue that tells the addict 'my drug is coming' will have the same potential effect. Eventually, the system stops even noticing the reward itself and just responds to cues that predict them. Other parts of the system besides the part mainly responsible for addiction start responding sluggishly when cues turn out to be disappointing. So the dopamine system chemically mimics the neuroeconomic model of the whole brain.*

A major barrier to the understanding of behavior by common sense is that everyday psychology misleadingly separates – or, at least, fails to strongly connect – the passive, perceptual aspect of reward and the active response to rewarding prospects. This partly stems from everyday psychology's overly direct connection of pleasure with reward; pleasure, as we have seen, is relatively passive and weak as a response motivator. But natural selection did not build massive learning systems for us so that we could *contemplate* complex attractions; in the brain, learning is for doing. Note also (a point to which we will return in more detail shortly) that thinking about something (as opposed to thinking about other possible things) is not (as in *Hamlet*) the opposite of acting. It is a kind of action, since it involves the brain's neurochemically and neurodynamically *doing something* (i.e., using resources for this rather than that). We should thus not think of reward learning as akin to a kind of *perception*, however much it has that *aspect*; the reward system is, fundamentally, a mechanism for allocating neural resources and prompting behavior. VTA and SNpc directly project to motor anterior cingulate and motor pre-frontal cortex. By this pathway, dopamine signals entrain motor habits that can proceed without conscious mediation after learning. This includes both 'bodily' behavioral habits and 'thought' habits in the sense of automatic associations we make when presented with familiar stimuli.¹⁹¹ The portentousness of this for addictive and compulsive phenomena, in which people feel that they have lost deliberate control, will be evident. Where the reward system is concerned, paying attention to something and doing something about it are more or less the same thing – it can't tell the difference between caring about (e.g.) gambling and getting busy gambling.

Non-technical synopsis: *The reward system closely fuses learning what to attend to and acting on what it attends to. It's a system for drawing the organism to something so as to act on it, without any thinking in between.*

¹⁹⁰ See additional evidence for this as reported by Martin-Soelch, C., Leenders, K., Chevally, A., Missimer, J., König, G., Magyar, S., Milno, A., & Schultz, W. (2001). Reward mechanisms in the brain and their role in dependence: evidence from neurophysiological and neuroimaging studies. *Brain Research Reviews* 36: 139-149.

¹⁹¹ Schroeder, T. (2004), *Three Faces of Desire*, Oxford: Oxford University Press reviews experiments from the 1990s that we can now, in light of what we've since learned about the dopaminergic system, recognize as showing this.

The most dramatic indication of the intimate relationship between the reward / dopaminergic system and motor behavior, and one that will turn out to be profoundly important to our understanding of AG, is the range of consequences attaching to impairment of the system's capacity to send adequate dopamine signals. Parkinson's disease is a degenerative disorder of the dopamine system as vested in substantia nigra. Its symptoms are loss of motor control (expressed in movement inability or difficulty [akinesia], tremors and rigidity), impaired cognition (difficulties with concentration, planning and learning) and motivation.¹⁹² Lesser compromises of the system due to lesions or chemical antagonists produce subsets of these same symptoms to varying degrees.

The way in which we have presented findings to this point may have misled the reader into thinking that neuroscience first arrived at a new understanding of reward in the brain, and then discovered by poking about in the deeper chemical details that dopamine is its agent. In fact, the dopaminergic pathways, and their effects on behavior, were isolated first – and some time ago – through animal experiments and the study of clinical populations. *Why*, exactly, motor control, learning, and motivation should have been linked through one central neurotransmitter was only gradually worked out, and fMRI investigations were crucial to the discovery process. In the course of this, a neuropsychologically adequate concept of reward finally, and only very recently indeed, snapped into view. Our claim to be able to understand addiction and AG rests essentially on this new enlightenment.

As recently as 2003 there was still disagreement on the interpretation of the findings. In the most complete survey of reward system science that had been done to that time, Berridge and Robinson¹⁹³ argued that dopaminergic response does not mediate reward learning in the sense of that phenomenon as it was then modeled (as “capacity to predict rewarding events based upon associative correlations”). Rather, they suggested, the system mediates “attribution of incentive salience to otherwise neutral events.” By this they mean that dopamine response to a stimulus is the basis for the organism focusing and maintaining attention (both perceptual and behavioral) on that stimulus. As Schroeder¹⁹⁴ explains “experiments in which the dopamine-releasing cells of the VTA projecting to the motor PFC were destroyed found that this impaired monkeys’ abilities to keep a prior intention in mind long enough to execute it after a delay.” Schroeder cites Durstewitz and colleagues¹⁹⁵ as suggesting “that the primary role of dopamine in the motor PFC is precisely to stabilize its goal-serving motor intentions against interference by other possible motor intentions that would be counter-productive. That is, information about expected reward is used to keep people and other animals like us focused upon our reward-directed intentions, to prevent us forgetting them or going off to do something else before carrying them out.”¹⁹⁶

¹⁹² Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron* 36: 241-263. p. 247.

¹⁹³ Berridge, K., & Robinson, T. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Research Reviews* 28: 309-369. (The quotation in the following sentence is from page 313.)

¹⁹⁴ Schroeder, T. (2004). *Three Faces of Desire*. Oxford: Oxford University Press (quotation from p. 116).

¹⁹⁵ Durstewitz, D., Kelc, M., & Gunturkun, O. (1999). A neurocomputational theory of the dopaminergic modulation of working memory functions. *Journal of Neuroscience* 19: 2807–2822.

¹⁹⁶ Schroeder, T. (2004). *Three Faces of Desire*. Oxford: Oxford University Press

Non-technical synopsis: *Two researchers argued a few years ago that the job of the dopamine system isn't to learn to predict reward. Instead, they argued, it's to pull something out of a background for attention and stay focused on it, not letting it slip back into the scenery again.*

The reason it was somewhat natural for Berridge and Robinson to frame this idea as a *rival* to the reward-learning model of the dopaminergic system stems from the fact that reward still had not (in 1998) been fully disassociated from pleasure. Their survey¹⁹⁷ devoted substantial critical attention to establishing that dopamine projections to the Nacc and neostriatum are not needed for “normal hedonic evaluations, for hedonic modulation, or for learned adjustments in hedonic value” – which is quite so. They noted that incentive salience attribution corresponds roughly to what behavioral economists had dubbed ‘decision utility’, defined as “the degree to which a goal is chosen or sought”. They recognized, of course, that decision utility – ‘wanting’ – and hedonic appreciation – liking – are related. Thus they suggested that once decision utility is associated with a stimulus then “on each subsequent encounter ... its capacity to support wanting is maintained or strengthened by associative reboosting of the incentive salience assigned to its representation. Reboosting happens when a wanted incentive is followed again by activation of hedonic liking”.

Non-technical synopsis: *The reason the two researchers mentioned above thought that dopamine wasn't for reward learning is that they knew it wasn't responding (much) to pleasure – and they hadn't quite realized how weakly related reward and pleasure are in the brain.*

In retrospect this critical quarrel – as is usually the case with apparent scientific disagreements – was really a step along the road to completing consensus on the conceptual clarification of ‘reward’ and its distinctions from related phenomena. “The compatibility of specific reward learning models,” declared Berridge and Robinson, “hinges on precisely what ... authors of reward learning hypotheses for dopamine function ... mean by ‘expectation of reward’”. If these reward learning hypotheses posit that dopamine mediates Pavlovian learning about *incentive value alone* (not stimulus-reward association per se, not hedonic value, not cognitive expectation of outcome) and that anticipatory neuronal responses to conditioned stimuli occur because those stimuli carry *incentive motivational* qualities, such as incentive salience, separately from mere association or from hedonic qualities, then those reward learning hypotheses are compatible with our incentive salience hypothesis ... If, however, dopamine neurons are posited to mediate any other sense of ‘expectation of reward’ ... then either the reward learning hypothesis or the incentive salience hypothesis must be wrong”.

Non-technical synopsis: *The researchers pointed out that we could, philosophically, separate reward from pleasure if we want to.*

Berridge's and Robinson's invitation to proponents of the reward-learning hypothesis to refine their concept of reward met, ultimately, with a favourable response. The first paper to state the mature model of the dopaminergic system, by McClure and colleagues,¹⁹⁸ essentially fused Berridge and Robinson's incentive salience hypothesis with Berns and Montague's predictor-valuation model (see section 3.1 above).

¹⁹⁷ Berridge, K., & Robinson, T. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Research Reviews* 28: 309-369.

¹⁹⁸ McClure, S., Daw, N., & Montague, P.R. (2003). A computational substrate for incentive salience. *Trends in Neuroscience* 26: 423-428.

All recent theories of the dopaminergic system, they pointed out, “agree on the singular principle that dopamine function is causally located between the identification of a potential future reward and the generation of action to pursue it”. However, the incentive salience hypothesis “leaves off after the identification of a goal.” What they mean by this is that it is a theory of the behavioral meaning of dopamine signals, but incorporates no account of learning except for the vague talk of ‘reboosting’ that Berridge and Robinson associate with it. This won’t do, since even if an initial dopamine response is taken to strongly bias selection of a goal, the system still must *choose* an action sequence, and choice in this context can’t be separated off from reward-valuation as expressed in relative motivational response. Temporal difference learning, as described in section 3.1, is precisely an algorithm for this, nudging the system towards actions that lead to better-than-predicted rewards. Here is where the predictor-valuation model can take up the slack: if the rate of dopamine neuron spiking encodes a prediction error signal, where incentive salience is then taken to be a function of the magnitudes of these errors, then the gap in Berridge and Robinson’s model is closed. “We propose,” say McClure *et al*, “that the concept of incentive salience is the expected future reward ... In addition, we propose that the role of dopamine in learning to attribute such expectations is to situations that are predictive of reward ... and in biasing action selection toward such situations ... serves as the formal counterpart to the ideas of Berridge and Robinson ... about the role of dopamine in attributing and using incentive salience”.

Non-technical synopsis: *The research team who created the neuroeconomic model of the reward system pointed out that a system for pulling out objects of attention from the background and keeping the agent focused on them couldn’t work if it wasn’t tightly integrated with decision making. But once those things get integrated that just is learning about what counts as rewarding for the system. So the neuroeconomic model makes a prediction about what the dopamine neurons are up to. And if that prediction is right, the model explains the dopamine system.*

As stated, here we have the current model of the reward system that has been confirmed by numerous experiments conducted over the past two years, several of which will be discussed below, and which will provide the basis for our later claim that AG will soon be a manageable health problem. The power of the model is nicely illustrated by application to a study – one directly relevant to gambling behavior – in which the interpreters failed to appreciate it. Zink and colleagues¹⁹⁹ compared skin conductance and **BOLD responses** (the name of the basic class of bloodflow variables measured under fMRI) of two groups of subjects. One group earned money by performing a task, while the other performed the same task but received equivalent expected quantities of money independently of task performance. As expected, significantly greater activity in the dopaminergic reward pathway (Nacc and caudate nucleus) was observed in the first group. Zink and colleagues interpret this as showing that dopaminergic activity mediates salience *rather than* expected future reward. Of course, they are quite correct about this if someone insists on interpreting ‘expected future reward’ as necessarily denoting the monetary value of the experimental outcomes. But this confuses what is rewarding to (as it were) the whole person with what is rewarding to *the reward system*. The point of McClure and coauthors’ mature version of the reward-predictor model is that what is rewarding to the reward system *is* incentive salience. Thus Zink’s ‘choice’ doesn’t arise: both interpretations are correct. For the first group of subjects in the Zink and co. experiment, the money was an *incentive* to action and *therefore* was salient – which is what ‘rewarding’ has turned out to *mean* in the brain.

¹⁹⁹ Zink, C., Pagnoni, G., Martin-Skurski, M., Chappelow, J. & Berns, G. (2004). Human striatal responses to monetary reward depend on saliency. *Neuron* 42: 509-717.

The second group (hedonically) enjoyed getting their money. Had they been previously unacquainted with what money can do, their subsequent use of it might *then* have been rewarding. But while choosing actions in the experimental task, their reward systems were uninterested in future hedonic prizes that were independent of subjects' actions. It should be obvious that this is significant for understanding AG, although it will take some time to draw out the details of exactly how it is.

With the model in hand, experimental work can proceed to refine it. For example, as McClure, Berns and Montague²⁰⁰ point out, the reward learning hypothesis does not distinguish between different possible dimensions of reward unpredictability (e.g.: What stimulus will occur next? When will the stimulus arrive? What quantity of stimulus will appear next?). They report an experiment that enabled them to isolate, under fMRI, predictions about what was expected by subjects within particular time frames. BOLD responses correlated with positive and negative errors against these temporally framed forecasts.²⁰¹ That is, subjects framed expectations that they'd receive rewards at certain times, and then their reward systems became most active at those times, because that's when the comparisons between what was expected and what was received were made. The reader might suppose that this is a problem for Rachlin's molar account of PG behavior as discussed in Chapter 1 section 7. If people's reward systems are sensitive to prediction errors against fixed time frames, then why would gamblers be insensitive to differences between shorter and longer strings of losses between wins? However, as McClure and his coauthors note, their experimental setup was extremely simple – subjects simply waited for alternating squirts of water and fruit juice – and this might have encouraged subjects to look for simpler temporal reference functions than would be applied to more typical situations, including situations set up in gambling. The study just described is, for now, merely an illustration of the kind of experiment we will see many more of in the course of gradually refining the reward-learning / predictor-valuation model. As will be clear in section 3.6, development of effective anti-AG pharmacological therapies need not wait on these refinements, though the efficiency of therapies will no doubt be improved by them.

Indeed, another recent refinement experiment is very good news for Rachlin's hypothesis about AG behavior. Nieuwenhuis and colleagues(2004)²⁰² performed fMRI examination of subjects involved in gambling scenarios where both the possible *kinds* of outcomes available relative to one another (large monetary gain, small gain, no gain, large loss, small loss, no loss) and the specific monetary magnitudes at stake were varied independently. They found that subjects' reward-system responses were much more sensitive to variations caused by context shifts than to variations in available absolute magnitudes.

²⁰⁰ McClure, S., Berns, G., & Montague, P.R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38: 339-346.

²⁰¹ Some supporting evidence from a study involving pigeons can be found in Kalenscher, T., Windmann, S., Diekamp, B., Rose, J., Güntürkün, O., & Colombo, M. (2005). Single units in the pigeon brain integrate reward amount and time-to-reward in an impulsive control task. *Current Biology* 15: 594-602.

²⁰² Nieuwenhuis, S., Heslenfeld, D., Alting von Geusau, N., Mars, R., Holroyd, C., & Yeung, N. (2004). Activity in human reward-sensitive brain areas is strongly context dependent. *NeuroImage* 25: 1302-1309.

For example, when subjects shifted from a scenario (i) in which there were possibilities of both no loss and a large loss to a scenario (ii) in which there was a possibility of no loss but no possibility of a large loss, their reward systems were more active in scenario (i), *and* – the important part here – this difference was greater than the difference between activation levels when they switched from scenarios where possible losses were large to scenarios where possible losses were small. Thus the reward system is more alert to *contrasts* in what *could* happen than in magnitudes of possible changes in wealth levels relative to the status quo. The exact logical mirror of this is a person who is more interested in a change in the excitement level of a gamble than in a change of odds in her favor. This is immensely interesting to students of AG, since the experimental subjects here were randomly sampled people, not PGs. Equally interesting, from the perspective of our interest in AG and in light of Rachlin's hypothesis, is the fact that some parts of the reward system differentially responded only to *binary* distinctions between outcomes (i.e., winning versus losing) and not to magnitude gradations at all. This is among the most positively suggestive neuroeconomic findings that a defender of Rachlin's hypothesis could hope for. The sort of experiment it suggests – performing comparative fMRI examinations of normal and PG subjects while they gamble and the length of strings of losses between wins is experimentally manipulated – would directly test Rachlin's theory of PG. If Rachlin's theory were confirmed by such an experiment, we'd then not only have new grounds for thinking that Rachlin is right, but the model of the reward system on which the experiment is based would then *explain* why he was right.

Non-technical synopsis: *Data suggest that people's reward systems are more attentive to changes in contrast between outcomes in gambles than they are to changes in the magnitudes of possible prizes in gambles. This is the very assumption Rachlin makes in constructing his model of the book-keeping mistake made by DGs (see Chapter 1, section 7). Thus if Rachlin is right, it seems that our model of the reward system explains why he is right.*

Of course the system as a whole does not *ignore* relative magnitudes of possible outcomes – that would be truly astonishing, and mysterious. Experimental work with rats reported by Tobler, Fiorello and Schultz²⁰³ suggests that dopaminergic responses adapt to changes in average magnitudes and to changes in magnitude variance (that is, changes in the distances between best and worst outcomes).²⁰⁴ This is consistent with the suggestion that the reward system is insensitive to magnitudes over shorter runs; indeed, Tobler and colleagues find this too. Pathological gamblers are attracted to larger stakes as time goes on, so they clearly aren't insensitive to prize magnitudes. But the experiment of Tobler and co. suggests that one source of their difficulties may lie in the fact that reward systems are relatively oblivious to win-loss magnitude ratios *within* gambling episodes.

Non-technical synopsis: *The reward system is not attentive to differences in magnitudes of outcomes within short time scales. This could also help to explain Rachlin's BE model of DG.*

²⁰³ Tobler, P., Fiorello, C. & Schultz, W. (2005), Adaptive coding of reward value by dopamine neurons, *Science* 307: 1642-1645.

²⁰⁴ Whole-system adaptation to reward magnitudes appears to implicate parts of the brain lying outside of the reward system as it is generally understood. See Glimcher, P., & Lau, B. (2005). Rethinking the thalamus. *Nature Neuroscience* 8: 983-984.

In general, the influence of ‘win salience’ vs. ‘loss salience’ contexts on the reward system seems to be profound. In earlier experimental work done over two decades ago, the great behavioral economists Kahneman and Tversky,²⁰⁵ found that in contexts in which losses are salient, subjects make riskier choices than in neutral contexts or those in which wins are salient. Recently, Gehring and Willoughby²⁰⁶ developed an experiment in which this result is replicated and pushed further. Randomly drawn subjects – not a selected sample of DGs – in a gambling task not only made riskier choices after losses but also made riskier choices the greater the magnitude of their preceding losses. Once again, the suggestive implications for AG in light of Rachlin’s hypothesis are clear. Gehring and Willoughby monitored their subjects using electroencephalogram, an older scanning technology, and found that the behavioral differences with respect to attraction to risk were correlated with changes (in what neuropsychologists call ‘event-related potentials’) in the ‘medial frontal negativity’, a region of medial frontal cortex near the anterior cingulate.²⁰⁷ They argue, consistently with the reward learning / predictor-valuation model, that this activity is coding the expected motivational impact of outcomes, not computing their expected value. Here, then, is another aspect of the trap waiting to lure gamblers – and not just pathological ones – to chase losses.

Non-technical synopsis: *The reward system chases losses, and chases them with greater attentiveness the larger they are. It seems our universal, inherited reward system is itself a DG! People who don’t behave like DGs must therefore be preventing their reward systems from exerting undue control over behavior. AGs may thus be people who, for some reason, can’t keep their reward systems in subordinate roles.*

This result is especially interesting in light of a contribution by Nakahara and colleagues.²⁰⁸ They had monkeys perform a task for reward under two conditions, first a condition in which reward probability given task performance was fixed, and then a condition in which reward probability changed over time as a function of the frequency of reward-delivery on previous trials. They measured response levels in dopamine neurons neuroelectrically. Their finding was that dopaminergic reward prediction takes account of learned context. They were only able to predict subject behavior using the temporal difference algorithm (see section 3.1 above) if they specifically added a context variable to the algorithm. This is a portentous theoretical modification, since it changes the learning rule in reward computing from one that responds to data as *it really is* and (given the results described immediately above) replaces it by one that responds to data as *biased by context*, and specifically, in gambling tasks, by recent history of winning or losing. Gambling losses might cause the primate reward system, with immediate effect, to revise the value of rewards upwards. Since the probability of winning will be estimated downwards with each additional loss, the difference between predicted value and received value of the betting action will grow as losses accumulate, enhancing the reinforcement effect of such wins as occur.

²⁰⁵ Kahneman, D. & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47:263-269.

²⁰⁶ Gehring, W. & Willoughby, A. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 295: 2279-2282.

²⁰⁷ This does not imply that the anterior cingulate, and therefore registration of relative levels of hedonic pleasure, are involved in the effect.

²⁰⁸ Nakahara, H., Itoh, H., Kawagoe, R., Takikawa, Y. & Hikosaka, O. (2004). Dopamine neurons can represent context-dependent prediction error. *Neuron* 41: 269-280.

Evidence for dynamics of this kind must increasingly prompt us to change our question from 'How are some people's reward systems primed to give rise to PG?' to 'How does anyone, given a normally functioning reward system, avoid PG?'

Non-technical synopsis: *In order to get the learning model of reward system response to predict some experimental data on monkeys, researchers had to adjust it so that loss chasing, and tendency to chase losses harder the longer it loses, is built directly into the mathematics. Increasingly, the research leads us to ask not 'Why are some people PGs?' but 'How is it that everyone is not a PG?' – just as the BE perspective described in Chapter 1 encouraged us to expect.*

The answer appears to lie in orbitofrontal cortex (OFC) and ventromedial prefrontal cortex (VMPFC) – or so we will suggest on the basis of a review of a variety of studies spread over all the remaining sections of this chapter. Both are sites of dopamine projection from VTA and SNpc. OFC has for some time been associated with control of impulsive behavior, while VMPFC seems to be the site where so-called 'somatic markers' – roughly, emotional significance tags – are attached to both perceptual and potential motor-response representations. (Note that conventional Western folk psychology has a well-entrenched tradition of thinking of emotions as *sources* of impulsive behavior. In light of neuroscience, it appears to be at least as true to say that emotions play essential roles in inhibiting and regulating such behavior.)

We will build up the case for this slowly, letting evidence to be reviewed speak for itself. In the remainder of this section we will focus just on studies that try to address the roles of VMPFC and OFC in molar reward responses. In later sections we will focus directly on the ways in which impairments to the functioning of these parts of the brain contribute to addiction and AG.

Such truth as there is to the traditional folk divide between cognition and emotion with respect to impulsive behavior may be suggested in an fMRI study by Pochon and colleagues.²⁰⁹ They monitored subjects using fMRI while they performed cognitively demanding tasks of varying levels of difficulty under two conditions: with and without monetary reward. In keeping with well-established findings, dorsolateral areas that implement working memory showed activation correlated with cognitive complexity of tasks. VMPFC was, as expected, more active in the reward-related condition. However, VMPFC activation *decreased* as tasks within this condition became harder. Pochon *et al* speculate that "[I]f one admits that one essential role of the positive reward is to act as behavioral reinforcer in simple tasks ... the same reward could induce a counterproductive emotional signal during the processing of a more demanding cognitive task by increasing the level of performance anxiety associated with the task" (p. 5673).

However, as most commercial games played by PGs involve only low cognitive complexity,²¹⁰ we are doubtful that this hypothesis is of direct relevance to our present subject.

²⁰⁹ Pochon, J., Levy, R., Fossati, P., Lehericy, S., Poline, J., Pillon, B., Le Bihan, D. & Dubois, B. (2002). The neural system that bridges reward and cognition in humans: An fMRI study. *PNAS* 99: 5669-5674.

²¹⁰ This point suggests an experiment. PGs should be examined under fMRI while the complexity of their gambling tasks varies. (It would be important to set novel gambling tasks in all conditions, to control for subjects' advance familiarity with games.)

The more interesting relationships where it is concerned lie *within* the reward system, between striatum and VMPFC / OFC. Here, the folk model of emotion and impulsivity breaks down entirely.

Non-technical synopsis: *Pre-frontal brain regions directly integrated into the reward system decline in activation as tasks become more difficult. To this extent, the traditional idea that cognition and emotion are opposing influences is borne out. But, we'll now see, this is as far as it goes; the traditional idea has more falsehood than truth in it.*

A substantial literature has accumulated showing that dysfunctions in somatic marking (i.e., impairment of emotional responses) predict poorer decision-making in cognitively demanding gambling tasks.²¹¹ On the other side, Shiv and colleagues²¹² showed, earlier this year, that in investment decision tasks with expected net positive returns ('expected' that is, in objective mathematical reality, not necessarily by the subjects) but possibilities of substantial losses, subjects with damage to emotional centres performed better than normal controls because (it seems) they experienced less fear of risk. This, of course, is not good news for emotionally impaired PGs, whose typical gambling scenarios do *not* feature expected net positive returns.

Non-technical synopsis: *Damage to emotional systems in the brain generally makes subjects worse in investment tasks. (That is, contrary to the traditional view, emotions usually help financial decision-making.) There are exceptions in cases where it's rational to take large risks, which emotion encourages people to flee. This isn't good news for emotionally impaired PGs – and many PGs are emotionally impaired by DSM criteria – because in commercial gambling scenarios odds are against gamblers so large risks are the opposite of rational.*

Some recent literature on the role of OFC in reward and motivation cites a study by Camille 'and colleagues'²¹³ which they interpret as showing that subjects with lesions to OFC make less advantageous gambling decisions than normal subjects because the former fail to anticipate and be motivated by the possibility of suffering from regret if they make a risky bet with negative expected return and lose. Unfortunately, we cannot rely on this study, because as Eagleman²¹⁴ shows, there was a discrepancy between the actual expected utilities of the gambles, on the one hand, and the expected utilities as the experimenters described them and as they presented them to subjects, on the other hand.

²¹¹ Bechara, A., Damasio, H., Tranel, D., & Damasio, A. (1997). Deciding advantageously before knowing the advantageous strategy. *Science* 275: 1293-1295; Rogers, R., Everitt, B., Baldacchino, A., Blackshaw, A., Swinson, R., Wynne, K., Baker, N., Hunter, J., Carthy, T., Booker, E., London, M., Deakin, J., Sahakian, B., & Robbins, T. (1999). Dissociable deficits in the decision-making cognition of chronic amphetamine abusers, opiate abusers, patients with focal damage to prefrontal cortex, and tryptophan-depleted normal volunteers: Evidence for monomaminergic mechanisms. *Neuropsychopharmacology* 20: 322-339.

²¹² Shiv, B., Loewenstein, G., Bechara, A., Damasio, H. & Damasio, A. (2005). Investment behavior and the negative side of emotion. *Psychological Science* 16: 435-439.

²¹³ Camille, N., Coricelli, G., Sallet, J., Pradet-Diehl, P., Duhamel, J.-R. & Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science* 304: 1167-1170.

²¹⁴ Eagleman, D. (2005). Comment on 'The involvement of the orbitofrontal cortex in the experience of regret'. *Science* 308: 1260b.

A final complication in the model of the reward system, before we turn to applications of the model in experiments on addicts and PGs, is suggested by studies that aim to discriminate between the impacts of VMPFC and OFC on discounting. fMRI work by Horn and colleagues²¹⁵ finds significant correlation between differences in impulsivity among subjects and differences in OFC activation levels, but not between impulsivity differences and variations in VMPFC levels. Yet, as we will see shortly, VMPFC impairment is correlated with addictive behavior (and therefore often gets associated with impulsivity in PG studies, insofar as PG is just assumed, following DSM-IV, to be an impulsive disorder or the direct consequence of such a disorder).

Non-technical synopsis: *A part of the brain that has been associated with addictive behavior has recently been found (in one experiment) not to be correlated in activation with impulsivity. But according to DSM-IV addictions are impulsive disorders. And what about all the evidence on this we reviewed in Chapter 2?*

What might explain this tension? Might VMPFC play a role in controlling addictive behavior that does *not* consist in restraining impulsivity? A novel experiment last year by Fellows and Farah²¹⁶ opens the window a crack on this question. As indicated in Chapters 1 and 2, we have here followed the behavioral economics literature in defining impulsivity narrowly, as being equivalent to steep discounting. However, as Fellows and Farah note, some behavior attributed to steep discounting might instead result from failure of agents to consider the implications of choices outside of a foreshortened future horizon. They refer to this construct variable as ‘future time perspective’. A behavioral economist would graph a short future time perspective as a kinked discount curve that increases in steepness after an inflection point. We are not aware of any BE models that experiment with such curves. Fellows and Farah compared four subject groups in a gambling task, and then measured the extent to which they spontaneously set long horizons when asked to imagine specific states of affairs ‘in the future’. The four groups were: (1) subjects with damage to VMPFC, (2) normal controls, (3) subjects with damage to dorsolateral frontal cortex (DLF), and (4) subjects with lesions away from the frontal lobes, but of comparable average size to lesions in groups (1) and (3). No differences were found with respect to discounting, or, within group (1), any correlation between *k*-values and sizes of lesions. On future time perspective, group (2) projected longer average futures than groups (3) and (4), and these two groups projected longer average futures than group (1) subjects. Thus VMPFC was uniquely implicated in generalized myopic thinking, as in other studies; but the evidence did not point to steep discounting as the mechanism.

Non-technical synopsis: *A recent study found a strong tendency for damage to the part of the brain traditionally associated with addiction to indeed predict low levels of attention to future prospects. However, these subjects were interpreted by the researchers as discounting normally insofar as they considered a given point in the future at all. Might BE be wrong, then, to explain addiction by reference to discounting?*

²¹⁵ Horn, N., Dolan, M., Elliott, R., Deakin, J. & Woodruff, P. (2003). Response inhibition and impulsivity: an fMRI study. *Neuropsychologia* 4: 1959-1966.

²¹⁶ Fellows, L. & Farah, M. (2004). Dissociable elements of human foresight: a role for the ventromedial frontal lobes in framing the future, but not in discounting future rewards. *Neuropsychologia* 43: 1214-1221.

This, at any rate, is Fellows and Farah's interpretation. As behavioral economists we find it a bit naïve, since maintaining that a subject's molar behavior (including her addictive behavior) is described in terms of some discount curve doesn't entail a claim that the curve in question is literally implemented in the brain – this is the molar / molecular distinction again, for which see Chapter 1, section 2. But this is a quibble,²¹⁷ not calling into question the significance of the study's results. If VMPFC projects reward estimates into future spaces that other parts of the reward system ignore, this will be relevant to understanding the behavioral implications of altered synaptic dispositions in cortex – for that is what appears to occur, as we will now see, in both substance addiction and AG.

3.3. Neuroscience of addiction

While neuroeconomists have worked to create functional models of the reward system, neuropsychologists studying addiction directly have tended to rely more immediately on neurochemistry. Ultimately, when we turn our focus squarely onto AG, we will want to deploy the *combined* power of the insights from the two literatures, which we think is greater than the sum of the parts. But this must first take us still deeper into the brain than we went in the previous section.

In Chapter 1 we pointed out that 'addiction' has often been used in loose and varying ways in behavioral science, so much so that behavioral economists have tended to avoid serious use of the concept. This is because economists group phenomena together by reference to their outward behavioral properties. Setting off 'addicts' as those who are allegedly chained to certain behavioral patterns by fear of withdrawal costs, as once was standard in psychology and psychiatry, seemed to rest on assuming the behavioral centrality of this property in advance of careful evidence for it. And then in fact the evidence, when it came in, went the other way, in the sense that most traditionally so-called 'addicts' turned out to be more capable of withdrawal than of avoiding relapse, despite their knowing that this implies facing the pain of withdrawal all over again at some point. If fear of withdrawal was the dominant and decisive factor for addiction, then this made no sense. And since no one had an alternative basis for drawing a sharp line between addicts and non-addicts, the concept went into eclipse among many of the scientists who study substance abuse and compulsive behavior.

Let us state clearly that we are not fussing here over whether we should or should not use the word 'addiction'. The issue, instead, is whether there is a fundamental difference, reflected in different appropriate treatment interventions and policies, between addicts and non-addicted people who sometimes have problems with substances or other targets of compulsive behavior. If there is, then the concept of addiction is a helpful one for scientific purposes. If there is no such difference, then it's best to drop the term, since it would encourage us to arbitrarily distinguish a subset of subjects.

The clinical community, by contrast with the behavioral science one, never stopped referring to 'addicts' as a special class of patients. We think that recent discoveries in neuroscience show that they were right. There *are*, it turns out, distinctive neural properties in a subset of people who abuse substances that are very strongly correlated with substance-abuse dispositions that are markedly more severe and intractable to various interventions than occur in any other people.

²¹⁷ It is particularly suspicious as a response in light of brand new results reported in Kable, J., & Glimcher, P. (forthcoming) Neural mechanisms for valuing immediate and delayed rewards in humans. *Society for Neuroscience Abstracts*. They provide evidence that molar (hyperbolic) curves fit molecular curves of reward system neurons very closely.

Both the behavioral properties and the neural properties of addicts have a very impressive degree of structure and pattern in common. The main features of this pattern are robust across particular personal and cultural circumstances. Then, as we'll see later, the behavioral patterns typical of addiction are highly responsive, in *detail*, to certain neuropharmacological interventions. Finally, these regularities aren't mysterious: they're unified by an elegant neurochemical model, to be described in this section, that meshes beautifully with the equally elegant model of the brain's computation of reward – the 'reward learning / predictor-valuation' model of neuroeconomics that was the subject of the previous two sections. In science, you can't get better grounds for thinking you've isolated a real phenomenon than having a cluster of robust causal regularities tied together by an exact quantitative model that generates measurable predictions you can test, and which keeps proving applicable to new phenomena. To which new phenomena have the neurochemical and neuroeconomic models turned out to apply? One answer is: PG. Pathological gamblers, it has turned out, are genuine addicts. As we will explain, this is itself *maximally* impressive confirming evidence for the models. And it explains why, in a survey on disordered gambling, we have spent so much time on the neurochemistry and neuroeconomics of reward, and will now devote equal attention to the neurochemistry and neuroeconomics of addiction.

Let us begin by quoting a representative clinical operationalization of addiction, formulated by a group of researchers in 1998 who had their eyes on the neuroscientific data that was beginning to accrue: "Addiction, also known as substance dependence ... is a chronically relapsing disorder that is characterized by three major elements: (1) compulsion to seek and take the drug, (2) loss of control in limiting intake, and (3) emergence of a negative emotional state (e.g., dysphoria, anxiety, irritability) when access to the drug is prevented".²¹⁸ Note that all of these aspects can be identified and induced in laboratory animals. As a result, when Koob and coauthors stated this operationalization, they already knew that it described a behavioral complex that was strongly correlated with unusual profiles in animals' dopaminergic systems. These systems are composed of two major projection pathways: one from SNpc to striatum and one from VTA to Nacc, frontal cortex and amygdala. Stimulant drugs such as cocaine and amphetamines inhibit reuptake of dopamine after it has facilitated synaptic transfer, so that it accumulates in Nacc. Opiates recruit additional mechanisms (mu-opioid receptors) in VTA and Nacc, while alcohol inhibits **GABA** (another neurotransmitter) neurons throughout the brain. Since GABA neurons inhibit dopaminergic neurons in VTA, the second-order result of GABAergic inhibition is the same flood of dopamine into Nacc that is characteristic of cocaine. Alcohol additionally promotes serotonin and mu-opioid receptors. Similarly, nicotine effects functioning in both the mesolimbic dopamine pathway and the opioid peptide circuit.

Non-technical synopsis: *Different substances to which people become addicted act on different brain chemicals and on different parts of the brain. So-called stimulant drugs, such as cocaine and amphetamines, act most directly on the reward system we have been discussing. But all act on it at least indirectly.*

Given what we have learned about the reward system, we should not be surprised that modified dopamine flow and concentrations are correlated with some of the familiar properties of addiction. Dopamine, after all, facilitates focus on salient reward predictors and primes the system for motor action in response to such predictors. A key part of addiction is clearly obsessive concentration on the target of abuse, coupled with a relentless tendency to pursue it in action.

²¹⁸ Koob, G., Paolo Sanna, P. & Bloom, F. (1998). Neuroscience of addiction. *Neuron* 21: 467-476.

But of course this is still a very long way short of explaining the molecular mechanisms of addiction. Why do dopamine concentrations pair with obsessive pursuit of certain *specific* rewards, viz., stimulant drugs, opiates, alcohol, nicotine ... and gambling? (That is, why don't people with elevated dopamine levels in Nacc just impulsively consume whatever happens to be to hand in good supply?) What is it about dopamine concentrations that leads to the characteristic *dynamics* of addiction, in which addicts become **sensitized** to their object of addiction – that is, show heightened behavioral effects to stimuli that predict the object²¹⁹ – and consume steadily more of it even while their hedonic pleasure from it drops?

As with all products of natural selection, the brain is not designed in the way that an engineer would assemble a machine. That is, there are no simple one-to-one relationships between neurotransmitters or other chemicals and styles of behavior. A first complication is that interesting behavior types result from *interactions* among different neurotransmitters (and other brain chemicals, such as **peptides**). In the case of addiction, the key interacting neurotransmitters, in addition to dopamine and GABA, are **glutamate** and **serotonin**. A second complication is that some neurotransmitters – e.g., glutamate – enhance synaptic activity while others – particularly GABA – inhibit it. And a third complication is that even for a given neurotransmitter within a single neuroanatomical region (e.g., dopamine in VMPFC) one finds different classes of receptors that behave differently from one another in the same chemical context. Thus, for example, in prefrontal cortex there are four kinds of dopamine receptors (referred to, sensibly enough, as D1, D2, D3 and D4 receptors).

Non-technical synopsis: *Processes in the brain are very complicated.*

We can best illustrate the extent of this complexity with an example. Dopamine interactions with GABA and glutamate clearly play the key role in causing sensitization in addicts.²²⁰ The specific ways in which they do this, however, varies in detail from substance to substance. So, for example, dopamine synapses with both glutamate and GABA neurons in VMPFC. But GABA (inter)neurons synapse on these same cells. Then whether dopamine excites or inhibits a given cell depends on whether that cell is or isn't also receiving GABA transmission.²²¹ (Indeed, it's also sensitive to the *neuroelectrical* state of the cell membrane.²²²) In general, whether dopamine release in VMPFC stimulates or inhibits GABA and glutamate release depends on whether it is taken up by D1 or by D2 receptors. In the same chemical context, D3 receptor activity seems to *oppose* D1 and D2 activity. And then remember that GABA release inhibits (among other things) glutamate neuron response.

²¹⁹ Note that sensitization is, with respect to some purely behavioral indicators, the opposite of 'tolerance', that is, diminished tendency of a given unit per interval of a addictive target to interfere with capacities for alternative behaviors. See Berke, J. & Hyman, S. (2000). Addiction, dopamine and the molecular mechanisms of memory. *Neuron* 25: 515-532.

²²⁰ Steketee, J. (2002). Neurotransmitter systems of the medial prefrontal cortex: potential role in sensitization to psychostimulants. *Brain Research Reviews* 42: 203-228.

²²¹ Gullledge, A. & Jaffe, D. (2001). Multiple effects of dopamine on on layer V pyramidal cell excitability in rat prefrontal cortex. *Journal of Neurophysiology* 86: 586-595; Yang, C., Seamans, J. & Gorelova, N. (1999). Developing a neuronal model for the pathophysiology of schizophrenia based on the nature of electrophysiological actions of dopamine in the prefrontal cortex. *Neuropsychopharmacology* 21: 161-194.

²²² See the work of Yang and colleagues cited in the preceding footnote.

Notice, then, that to determine a prediction of whether a given boost of dopamine will produce a net increase or decrease in glutamate activation *just in this little bit of the circuit* will require both exact knowledge of which receptors are triggered and quantitative measurement of relative, dynamic, neurotransmitter levels at the start of the process. We italicized a few words in the previous sentence to emphasize that the example here concerns a relatively *simple* relationship within a complex macro-level circuit that involves distinct neuroanatomical and neuropsychological systems, multiple kinds of glutamate and GABA receptors, at least three other neurotransmitters in supporting roles (in the case of serotonin, perhaps a major inhibitory role), plus neuroelectrical modulation.

Non-technical synopsis: *Even very small parts of the processes by which the brain generates and modifies behavior are very, very complicated.*

The lay reader might be dizzy now – our point was to make her so – and perhaps also a bit dismayed. In terms of physical articulation, the brain is the single most complex functional system studied in all of science. If understanding its responses to stimuli requires, for each stimulus, a complete molecular pathway model, then it seems that all neuroscientific explanations will be hard to achieve, will be bogglingly complicated, and will be difficult to integrate into generalizable knowledge – or, where pathologies are concerned, general treatment strategies. For example, we have several times made reference to the effect on the brain of ‘stimulant drugs’, citing as examples the standard cases, cocaine and amphetamines. They indeed act in the same way up to the point of their influence on VMPFC. But Steketee²²³ shows in a review of research that, in fact, the ways in which they influence VMPFC, and then influence VMPFC’s effect on subcortical regions in processes of sensitization, are quite different.²²⁴

Does this imply that we’ll want different pharmacological treatments for cocaine and amphetamine addiction? Well, yes and no. As we’ll see shortly, these drugs have enough in common in the way they act that we can already synthesize a single sort of agent that works powerfully to reduce their addictive hold. To the extent that their neurochemical actions are different in the way Steketee indicates, this therapy will likely be considered crude after further research leads to a next generation of drugs; we will ultimately have one improved therapy for cocaine and another for amphetamines. At present, pilot studies are testing drugs developed for relief of cocaine addiction on pathological gamblers; while there is good reason to think these drugs will make a difference, it will be astonishing if this turns out to be our *ultimate* best response to AG.

The underlying point here is that the scientific study of a phenomenon never reaches a final resting point. At any given time we will know just so much about specific addictions, and there will be more we’re about to know still waiting around the corner. Thus for all addictions our scientifically based treatment strategies will go on indefinitely improving with our understanding. At present, we can rank our knowledge of the neurochemistry of the major substances of abuse, from greatest to least, roughly as follows:

- Cocaine
- Amphetamines

²²³ Steketee, J. (2002). Neurotransmitter systems of the medial prefrontal cortex: potential role in sensitization to psychostimulants. *Brain Research Reviews* 42: 203-228.

²²⁴ In fact, they also differ in the ways they increase dopamine concentrations in Nacc to begin with. Cocaine blocks reuptake and amphetamines stimulate efflux from synapses.

- Alcohol
- Nicotine
- Opiates

If we were writing a survey about cocaine addiction and treatment, it would be incumbent on us to dig down into quite fine levels of neurochemical detail, just because a lot is known at this level about cocaine. By contrast, if we were writing a survey on opiate addiction and treatment for lay readers, we would have to content ourselves with very general remarks because there are simply too many possible molecular accounts among which data presently don't decide. We know that opiate response is sufficiently distinctive that we certainly cannot just import principles from our understanding of cocaine to it.

Now, the contention about gambling that we'll defend is that it has an interesting status in terms of the list above. A very few years ago, it would have belonged at the bottom of the list. It was an open question, indeed, whether there even is such a phenomenon as 'addiction' to gambling in the sense of a distinctive type of neurochemical and neurodynamic processes associated with severe and chronic disordered gambling behavior. But now we have very strong evidence that not only is there such a thing as gambling addiction but that it resembles the better understood substance addictions, those at the top of the list above, more than it does those at the bottom. In fact, we'll present some reasons for thinking that addiction to gambling may be *the* most basic and fundamental form of addiction, the form around which can be built a general template for addiction models; models of targets with more specific and complex causal effects, such as heroin, will then be (considerable) complications to this template. We can summarize this claim as follows: *in studying addictive gambling we come as close as possible to studying the generic form of addiction.*

In working up to this claim, the first thing we need to do is indicate the *general* properties of stimulant drug addiction, abstracting from details that *separate* special pathways among the stimulant drugs. In reviewing leading data, we will add explicit notes drawing attention to aspects that are suggestive where AG is concerned. Though we will discuss some data on subjects addicted to other substances, especially nicotine, we will confine ourselves to aspects of these data that reflect on neurochemical patterns that all addictions have in common, namely, those directly associated with learning in the dopaminergic system.

Enhanced dopamine release into striatum contributes to the promotion of action from representation. Enhanced dopamine release into Nacc, specifically, promotes focus, attention and learning in accordance with the reward learning / predictor-valuation model. This provokes urges to exploratory behavior centred on the target of attention. Stimulant drugs act directly to increase dopamine in Nacc, thus causing effects that mimic behavioral response to novel stimuli.²²⁵

²²⁵ Chambers, R.A., Taylor, J. & Potenza, M. (2003). Developmental neurocircuitry of motivation in adolescence: A critical period of addiction vulnerability. *American Journal of Psychiatry* 160: 1041-1052.

In addition, dopamine activity acting on D3 receptors interacts with glutamate and GABAergic neurons in *both* Nacc and in VMPFC to alter the firing potentials of synapses in subcortical regions and thereby causes behavioral sensitization – that is heightened motor response to learned cues for the rewards associated with the substance.²²⁶ In particular, it causes so-called ‘long-term potentiation’ (LTP) at synapses, an implementation of memory. We know that D3 receptors in striatum are particularly coupled to movement control, because – something that will later emerge as important in research on PG – animals given agents that block D3 receptors develop Parkinson’s symptoms.²²⁷ Serotonin (5-HT) release inhibits excitatory effects of dopamine on subcortical areas and so low serotonin is correlated with both addiction learning and with probability of relapse.

Non-technical synopsis: *All addictive drugs either directly or indirectly flood a particular part of the brain with dopamine. This causes the reward system to respond as if a novel, unpredictable stimulus is at hand. In addition, addictive drugs physically modify memory so that the brain automatically begins implementing certain action patterns when it picks up a cue that predicts the drug. The part of memory that is effected is the same part that is effected, in the opposite way, by Parkinson’s disease.*

It is well known that younger people, especially adolescents, are much more prone to develop addictions, and to be less able to avoid relapse during recovery attempts, than older people. The large majority of addictions begin in adolescence, at least in the minimal sense that their adult occurrences can be predicted from adolescent behavior. It is therefore indirect confirming evidence for the general model above that, as Chambers, Taylor and Potenza²²⁸ describe, adolescent mammals have heightened dopaminergic responses and diminished inhibitory 5-HT systems. Furthermore, prefrontal cortex, which plays an inhibitory role in the development of addiction and which dopaminergic action must hijack during potentiation for addiction to occur, is not yet fully mature in adolescent humans. Their brains adapt faster and more permanently to stimuli in general. Thus it is not surprising that since addiction is a variety of adaptive learning, young people are more vulnerable to it.²²⁹

Non-technical synopsis: *Brains of adolescents are more easily captured by addictive drugs than brains of adults, because the adaptive part of the adolescent brain is more receptive, and the part of the brain that interferes with adaptation isn’t yet mature in adolescents.*

There is direct evidence that the neural learning of addiction requires active exploration by subjects. That is, brains don’t adapt to drugs by simply ‘soaking them up’; they adapt by forming new patterns of active response to make receipt of drugs more likely. **Plasticity** is the term for the disposition of a learning unit (anything from the whole brain to a single synapse) to adapt to new input.

-
- ²²⁶ Everitt, B., Dickinson, A. & Robbins, T. (2001). The neuropsychological basis of addictive behavior. *Brain Research Reviews* 36: 129-138.
- ²²⁷ Baik, J., Picetti, R., Saiardi, A., Thirlet, G., Dierich, A., Depaulis, A., Le Meur, M. & Borrelli, E. (1995). Parkinson-like locomotor impairment in mice lacking dopamine D2 receptors. *Nature* 377: 424-428.
- ²²⁸ Chambers, R.A., Taylor, J. & Potenza, M. (2003). Developmental neurocircuitry of motivation in adolescence: A critical period of addiction vulnerability. *American Journal of Psychiatry* 160: 1041-1052.
- ²²⁹ Gray, P. B. (2004). Evolutionary and Cross-Cultural Perspectives on Gambling. *Journal of Gambling Studies*, 20(4), 347-371.

At excitatory (glutamate) synapses implicated in addiction learning, a measure of plasticity is the ratio of two types of glutamate receptors, AMPA to NMDA. Dumont and colleagues²³⁰ performed an experiment in 2005 which revealed that self-stimulation by means of cocaine and food pellets by rats changes this ratio, whereas passive administration of the same quantities of these rewards doesn't change it. This is consistent with some venerable puzzles about addiction, such as the fact that hospital patients given high doses of morphine for extended periods don't generally develop addiction. The result vividly underscores the way in which the reward learning / predictor-valuation model, as an account of the learning of salient *relationships* between environmental cues and actions, captures addiction. This feature will emerge as central in the account of AG.

Non-technical synopsis: *Brains learn to be addicted not just by receiving drugs but by developing active patterns aimed at recruiting them. This is why patients given morphine in hospital, who don't decide to take it themselves or do anything to solicit it, seldom become addicts despite often receiving regular addiction-sized doses. The brain's chasing of addictive targets is part of what it is to become addicted.*

Bechara and Martin²³¹ found that human substance addicts suffer from measurable impairment of working memory relative to controls. This probably does not reflect any physical compromise of brain areas, but is instead a function of the focusing effect of dopaminergic learning, which inhibits attention to anything other than the addictive target. A number of the performance differences between substance abusers and controls on the Iowa Gambling Task as reviewed in Chapter 2 may be explained by this aspect of the neuropsychology of addiction. It may be especially relevant to AG in light of the well-known fact that PGs' behavior while gambling is much less sensitive than other people's to shifts of odds against them. The greater motivation of PGs to gamble may be complemented by greater difficulty they have in cognitively processing the implications of changing odds. This evidence is supported in another recent study by Garavan and Stout.²³²

In their experiment, human addicts showed poorer than normal performance at cognitive learning (in attention-switching, Stroop²³³ and inhibitory control tasks) based on positive and negative feedback from other people. Interestingly, during processing of these cognitive tasks addicts under fMRI showed strongly enhanced activity in anterior cingulate cortex, another direct site of projection from VTA and SNpc, relative to controls.

-
- ²³⁰ Dumont, E., Mark, G., Mader, S. & Williams, J. (2005). Self-administration enhances excitatory synaptic transmission in the bed nucleus of the stria terminalis. *Nature Neuroscience* 8: 413-414.
- ²³¹ Bechara, A. & Martin, E. (2004). Impaired decision making related to working memory deficits in individuals with substance addictions. *Neuropsychology* 18: 152-162.
- ²³² Garavan, H. & Stout, J. (2005). Neurocognitive insights into substance abuse. *Trends in Cognitive Sciences* 9: 195-201.
- ²³³ 'Stroop' tasks measure the difference in how long subjects take to respond to stimuli that they are all supposed to treat in the same way, but which differ in a way chosen by the experimenter. For example, subjects might be asked to read aloud the words presented on a screen, and the experimenters discover that subjects take slightly longer to read the words associated with suffering than neutral words. Delay in reading is an index of more internal processing activity having taken place. Data from Stroop tasks can help neuroscientists work out what they are looking for.

Addicts' cognitive performance improves when they take a shot of their drug, which is consistent with the reward learning / predictor-valuation model: drug reward temporarily reduces the distance between predicted reward state and actual reward state, and so dampens the mesolimbic noise that interferes with cognitive processing. Smokers are all aware of the greater pressure they feel to reach for their drug while trying to concentrate. Of direct relevance to PG, in the Garavan and Stout experiment, hypoactivity of anterior cingulate in addicts is correlated with lower response to losses during cognitive learning.

Non-technical synopsis: *Because addiction concentrates an addict's attention on her drug, addicts behave as if they suffered from impaired memory and cognition. Some do, but this isn't always or necessarily the explanation; in the first place, it's that their resources are otherwise engaged (just as BE predicted). This may be related to PGs' reduced sensitivity to odds that shift against them during games. Addicts' attentiveness to other stimuli increases when they get a shot of their drug, as neuroeconomics predicts, because this decreases the difference between actual reward and expected reward and so quiets the reward system (briefly).*

A suggestive study of the relationship between addiction and attentional gating is reported by Martin-Soelch and colleagues.²³⁴ They conducted a PET neuroimaging study involving 4 groups performing cognitive tasks for monetary and non-monetary rewards: (1) normal controls, (2) smokers, (3) opiate addicts, (4) Parkinson's sufferers. Group (1) showed diverse neural responses to both sorts of rewards, in all the brain regions that our knowledge of the reward system would lead us to expect. However, smokers showed response only in OFC, right cingulate gyrus and left midbrain – nothing in striatum (or thalamus). Opiate addicts showed less response than controls in dopaminergic areas to all rewards, and no response in dopaminergic areas to non-monetary rewards. Group (4) matched group (1) except with respect to their lack of OFC and striatal responses. Thus we have more illustration of the extent to which addiction focuses the reward system to the exclusion of targets other than the object of addiction. It is interesting, as the researchers note, that nicotine appeared to be a more powerful agent in this regard than opiates. If this surprises the lay reader, then she should use the occasion to reflect on *why* she is surprised. Our everyday idea of 'impulsiveness' incorporates reference to specific *actions* that reflect impaired judgment. In this respect, smoking will seem the least dangerous of addictions (aside from its impact on organs other than the brain). However, the relevant concept of impulsiveness in the neurochemistry of addiction is the neuroeconomic, not the everyday one. The impact of nicotine on smokers' *molar* patterns of judgment appears to be relatively profound. This is almost certainly implicated in the notoriously high rates of relapse for people trying to stop smoking. A similar confusion between everyday and neuroscientifically grounded psychological concepts probably contributes to some of the prevailing skepticism about gambling as a target for genuine *addiction*. The facts about nicotine, however, show that there is no necessary relationship between addictive impulsiveness and episodic mania.

Non-technical synopsis: *There is no necessary connection, as the popular understanding of drug influence often assumes, between socially eccentric behavior (intoxication) and the extent to which drugs change cognitive attention. Nicotine doesn't cause people to act wildly, but experiments show that it gates out incoming information from other brain areas even more effectively than opiates.*

²³⁴ Martin-Soelch, C., Leenders, K., Chevally, A., Missimer, J., König, G., Magyar, S., Milno, A. & Schultz, W. (2001). Reward mechanisms in the brain and their role in dependence: evidence from neurophysiological and neuroimaging studies. *Brain Research Reviews* 36: 139-149.

(This is why people smoke to improve concentration.) Thus, to the extent that people are skeptical about gambling addiction because gamblers remain (relatively) lucid, this has no scientific weight as a consideration.

Another important study showing effects of dopaminergic attentional gating in addicts is due to Garavan and colleagues.²³⁵ They found that cocaine addicts show significantly less activation in the projection from the VMPFC to the Nacc than normal controls on being shown sexually evocative pictures. Perhaps this is why addicts perform poorly on the Iowa Gambling Task relative to non-addicts, as discussed in Chapter 2.

Non-technical synopsis: *More evidence for the idea that addictive drugs reduce attention to other stimuli is that the reward systems of cocaine addicts show less reaction to sexy pictures than reward systems of normal controls.*

Kalivas and colleagues²³⁶ report 2005 work that refines our understanding of behavioral attentional gating in addicts by separating its two underlying neurochemical substrates. Activation patterns in recovered rat cocaine addicts presented with cocaine-associated cues suggests that dopamine release in the projection from the PFC to the Nacc is responsible for strongly cueing the learned expectation of reward in response to the cues, while adaptations in glutamate synapses in Nacc reduce plasticity and impair the animals' abilities to learn to respond to alternative rewards. Since the drug in this case is cocaine, the closest drug model to gambling, our interest in AG should lead us to attend to some molecular details. Kalivas *et al* appeal to a neurochemical model of the dopamine action due to Siemans and Yang.²³⁷ On this model, dopamine action gives rise to two possible states in VMPFC depending on whether influence at D1 or D2 receptors predominates. If D2 reception predominates, then multiple excitatory inputs promote VMPFC output to Nacc. If D1 reception predominates, then all signals below a high threshold are inhibited. In cocaine withdrawal, protein signaling to D2 receptors is reduced, thus inducing the animal to seek stimuli that can clear the high D1 threshold – namely, stimuli associated with cocaine. Here, then, is a neurochemical model of the mechanism by which withdrawal gives rise to cravings.

Non-technical synopsis: *The way in which cocaine produces concentration of attention on itself in the reward system involves two interacting molecular mechanisms. One lowers the gate for cocaine and another pushes up the gate for anything else. If the system doesn't receive cocaine, the lowered gate stops receiving input and the organism goes looking for something that causes a strong enough dopamine boost to clear the heightened gate – which, in this case, will be cocaine. So this is the neurochemical explanation of cravings during withdrawal.*

Further details have come to light on the way in which dopamine and glutamate neurons complement one another in subserving reward learning and in generating addiction.²³⁸

-
- ²³⁵ Garavan, H., Pankiewicz, J., Bloom, A., Cho, J., Sperry, L., Ross, T., Salmeron, B., Risinger, R., Kelley, D. & Stein, E. (2000). Cue-induced cocaine craving: Neuroanatomical specificity for drug users and drug stimuli. *American Journal of Psychiatry* 157: 1789-1798.
- ²³⁶ Kalivas, P., Volkow, N. & Seamans, J. (2005). Unmanageable motivation in addiction: A pathology in prefrontal-accumbens glutamate transmission. *Neuron* 45: 647-650.
- ²³⁷ Seamans, J. & Yang, C. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Progress in Neurobiology* 74: 1-57.
- ²³⁸ Kelley, A. (2004). Memory and addiction: shared neural circuitry and molecular mechanisms. *Neuron* 44: 161-179.

Cells in Nacc and VMPFC on which their dopamine and glutamate neurons jointly synapse act as coincidence detectors in associative learning; essentially, an animal or person is motivated to act when dopamine and glutamate signals agree with one another in sending positive signals. Glutamate responds to specific sensory information, and ‘teaches’ the system to respond to new information with a similar profile (by, like any other neurotransmitter system, modifying synaptic potentials). It thereby lays down episodic memories that are guides to subsequent actions. Dopamine neurons, by contrast, respond to *global* saliences, i.e., overall states of the world that suggest reward prospects. We now understand the basic molecular neurochemistry of dopaminergic influence. In ventral striatum and VMPFC neurons are normally held quiet by very negative resting membrane potentials (called ‘K+’ currents) unless pushed into de-polarized states by inflow from cerebral cortex and thalamus. When they’re depolarized dopamine excites them; when they’re polarized it dampens them further. Thus motivation begins with cognitive judgments (which may reassure a reader who was beginning to feel like a robot or a rat from reading this survey). However, the reward system is designed to both amplify these initial judgments and make them difficult to behaviorally revise; natural selection decided, as it were, that we should not waste precious cognitive resources reconsidering judgments. Here are the roots of addiction, which we can now see are products of two mechanisms. First dopamine takes an initial positive response to a stimulus and makes it a self-reinforcing obsession. (We will return to the question of why it does so for some rewards and not others later.) Meanwhile glutamate ‘stamps in’ the dopaminergic response by modifying synapses to be alert to distinctive sensory predictors. Thus the smoker cannot bear to finish a meal without lighting up, the drinker cannot pass the bar without going in, and the gambler experiences a compulsive rush when she hears jangling coins.

Non-technical synopsis: *The basic dynamics of the first stage of reward-system adaptation to targets of addiction works as follows. First the person encounters a reward from an addictive target, either on the basis of a cognitive judgment that it’s interesting or looks fun or is a good means to some other end, or by simply ‘bumping into’ it. The dopamine system then amplifies the initial judgment that the stimulus is rewarding. A second neurochemical system, the glutamate system, then modifies memory and learning pathways so that the dopamine system becomes yet more alert to cues that predict the reward. As this cycle keeps strengthening through feedback, ever more features of the person’s environment will come to be associated with the drug or other addictive target and will trigger cravings for it.*

Powerful though the reward system is in regulating behavior, however, it is not in charge. Though dopamine concentration in Nacc is necessary for addiction, it is not sufficient for it; after all, stimulant drugs produce high concentrations of dopamine in Nacc of non-addicts, and most people who are exposed to addictive substances or gambling do not become addicts. As noted above, cognitive judgment plays an initiating role in motivation. But it furthermore plays a regulative role at all times – except when addiction neurochemically subverts it. Goldstein and Volkow²³⁹ review a now substantial body of evidence that addictive targets do more than hijack the mesolimbic reward system; they also impair inhibition of subcortical responses by prefrontal cortex (PFC) circuits. One of the ways that some drugs do this is as directly as possible: stimulants, alcohol and opiates physically destroy some frontal circuitry.

²³⁹ Goldstein, R. & Volkow, N. (2002). Drug addiction and its underlying neurobiological basis: neuroimaging evidence for the involvement of the prefrontal cortex. *American Journal of Psychiatry* 159: 1642-1652.

More subtly, consumption of addictive substances appears to cause reformation of dendritic links (the electrical, as opposed to chemically mediated, connections between neurons; see section 3.2) as a result of which normal cortical inhibition of the amygdala is reduced.²⁴⁰ In effect, addiction not only hijacks the limbic system, but sabotages the systems that might check its influence.

Adding this evidence to that which we have reviewed on the dopamine-glutamate mechanism in Nacc, Goldstein and Volkow construct an integrated model of addiction which they call ‘impaired response inhibition and salience attribution’ (I-RISA). This model has become the standard one in the neuroscience of addiction. It links the neural processes underlying four distinct aspects of addiction to substances: intoxication, craving, compulsive drug administration and withdrawal. The main distinctive general feature of intoxication is dopamine concentration in Nacc. Craving is then the learned association between the drug reward and various stimulus contingencies, based on modified synaptic potentials in amygdala, hippocampus, thalamus, anterior cingulate and OFC. Compulsive drug administration is then the consequence of the mesolimbic reward system that has learned to treat the addictive target as an overwhelmingly salient reward acting semi-autonomously in controlling behavior. Strong dopaminergic and serotonergic responses then support withdrawal, symptoms of which include dysphoria (sadness), anhedonia (hedonic discomfort), impaired cognition, and irritability. As we have argued, fear of withdrawal is almost certainly not the main factor in maintaining addiction. Instead, as a result of the re-sculpting of prefrontal circuits that reduces cortical inhibition of the limbic system, the cravings induced by the addiction-trained reward system, against which alternative rewards have difficulty competing, are often sufficient to provoke relapse after periods of abstinence. Because of the way in which the reward system learns, cravings are induced by any of the cues with which the target of addiction was associated during learning.²⁴¹

Non-technical synopsis: *The full neurochemical model of drug addiction is known as ‘I-RISA’. It integrates intoxication, craving, compulsive drug-taking and withdrawal. Intoxication is the dopamine flood into a part of the brain, interfering with cognition and emotion. Craving results from the way in which the reward system learns to associate more and more features of the environment with the drug as a predictor of it. Because this system is a direct producer of action, it now causes the person to compulsively consume the drug. The thresholds any other rewards need to stimulate the system are raised, thus causing sadness, hedonic pain, and cognitive and emotional discomfort if the person stops consuming the drug – this is withdrawal. Circuits in the newer, more cognitive, part of the brain that prevent the reward system from always getting its own way are re-wired in such a way as to subvert this function. Thus the person is likely to relapse after stopping consumption, even when her withdrawal symptoms are mild.*

²⁴⁰ For evidence and details see Robinson, I., Gorny, G., Milton, E. & Kolb, B. (2001), Cocaine self-administration alters the morphology dendrites and dendritic spines in the nucleus accumbens and neurocortex, *Synapse* 39: 257-266; Rosenkranz, J. & Grace, A. (2001), Dopamine attenuates prefrontal cortical suppression of sensory inputs to the basolateral amygdala of rats, *Journal of Neuroscience* 21: 4090-4103; Miller, E. & Cohen, J. (2001), An integrative theory of prefrontal cortex function, *Annual Review of Neuroscience* 24: 167-202.

²⁴¹ See Chiamulera, C. (2004), Cue reactivity in nicotine and tobacco dependence: a “multiple-action” model of nicotine as a primary reinforcement and as an enhancer of the effects of smoking-associated stimuli, *Brain Research Reviews* 48: 74-97, for a detailed application of this model to the neural substrates of nicotine addiction and relapse in recovery attempts. In keeping with the reward learning / predictor-valuation model, Chiamulera refers to cravings as “failed expectations”.

Most aspects of the I-RISA model are anticipated in a general account of the neural learning and maintenance of addiction given by Everitt, Dickinson and Robbins.²⁴² They add further details on the functional neuroanatomy of the process that we have not yet mentioned. In particular, they distinguish between the Nacc ‘shell’ and other Nacc subregions. Animals with lesions to non-shell subregions of Nacc show impaired learning of addictions; this does not apply to Nacc shell lesions. Addictive learning is also impaired by lesions in the central nucleus, but not the basolateral area, of the amygdala. This is initially puzzling, since the basolateral amygdala, but not the amygdala central nucleus, projects to Nacc. The answer Everitt and colleagues suggest to this puzzle is that amygdala central nucleus regulates the dopamine neurons of VTA, thereby affecting Nacc response. The Nacc shell regulates its own dopamine innervation via projections to the VTA and the Nacc core. We consider the neuropharmacological implications of these findings in section 3.6.

Non-technical synopsis: *This paragraph describes some further details on the pathways of addictive learning, and so cannot be described non-technically.*

What the I-RISA model does *not* incorporate is a theory of what causes addiction in the first place. This question decomposes into two further ones: (i) What distinguishes things that can be targets of addiction from things that aren’t? (ii) What distinguishes those who become addicts from those who don’t?

On the second question, genetic factors tell at least part of the story. Laakso and colleagues²⁴³ report early work on the search for a genetic basis of vulnerability to addiction. The obvious first place to look is at genes that regulate the dopaminergic system. Unfortunately, there are many of these. A basic technique for trying to isolate culprits for addiction is by selectively ‘knocking out’ dopamine-regulating genes from animals and then searching for differences in susceptibility to addictive substances that correlate with these knockouts. A main result so far is that “[o]f the many knockout mice tested to date in paradigms of drug addiction, the mGluR5 mutant mouse is the only line that fails to self-administer psychostimulants” (p. 221). For an alternative technique for isolation of addiction-related genes, based on breeding of so-called ‘congenic’ animals, see Wise (2000),²⁴⁴ p. 31. A third approach is to rear twin animals and selectively alter genes or suppress gene activation in some twins but not others. We return to the more specific question of possible genetically based predispositions for AG later.

Stress from the environment has been clearly identified, both in folk experience and by science, as causally significant for addiction. In the brain, stress is manifested when hypothalamic corticotropin-releasing factor (CRF) releases adrenocorticotrophic hormone (ACTH) from the pituitary-adrenal system, which in turn releases glucocorticoids. Stress also activates the sympathetic nervous system and emotional systems.

²⁴² Everitt, B., Dickinson, A. & Robbins, T. (2001). The neuropsychological basis of addictive behavior. *Brain Research Reviews* 36: 129-138.

²⁴³ Laakso, A., Mohn, A., Gainetdinov, R. & Caron, M. (2002). Experimental genetic approaches to addiction. *Neuron* 36: 213-228.

²⁴⁴ Wise, R. (2000). Addiction becomes a brain disease. *Neuron* 26: 27-33.

When stress persists through failure of attempts to relieve it, it recruits a range of neural systems and can ultimately affect most of the brain.²⁴⁵ Even mild stress has direct implications for limbic activity, so it is no surprise that stress and reward system pathologies are closely linked. Obviously, however, in the overwhelming majority of cases where people are stressed, including cases where they are stressed for prolonged periods, no addiction results. Nor does stress seem to be a necessary condition for addiction onset, since drugs that relieve stress are themselves potentially addictive. It will not surprise any researchers if addiction's developmental etiology turns out to lie in complex interactions between genetically based predispositions and contingent sources of stress from the environment, perhaps with a small subset of people whose genetics are such that they are almost sure to develop addictions regardless of their life circumstances. While developing an effective medical *preventative strategy* against addiction would require us to know more than we do now about its ultimate neurochemical causes, development of effective medical *interventions* in addiction need not wait upon such knowledge; our rapidly expanding knowledge of proximate mechanisms as reviewed in this section is taking us a long way.

3.4. Gambling as elementary reward system manipulation

The other open question above, why certain reward sources are regular targets of addiction and others aren't, is fundamental to our concern with AG. Tradition has taken drugs (including alcohol) to be the core addictive targets for a number of reasons. First, all of the leading ones except nicotine produce vivid episodic 'highs' in which judgment is impaired and hedonic pleasure is enhanced (at least for a time). This means that for most drugs there is a strong distinction between sober and intoxicated states. Furthermore, it is straightforward to comparatively measure any given person's quantity and rate of consumption of drugs. Folk wisdom therefore doesn't usually need to distinguish between addiction and very heavy or frequent use. Given social disapproval of intoxication, this is a reasonable approximation. A person who is very often drunk probably *is* addicted to alcohol. The majority of other recreational drugs are illegal, so once again the rough assumption that regular users are addicts is sustained by high imposed costs: a few very wealthy people aside, who but an addict would keep anteing the cash and running the risks to maintain a constant supply of cocaine?²⁴⁶

In the context of this popular wisdom, it is not surprising that talk of addiction to non-drug rewards is often regarded as metaphorical. It is common for people to be referred to as 'addicted' to television or sex or shopping, but all that is typically meant by this is that the activity in question consumes an unusual proportion of the 'addict's' time – a proportion the attributor regards as socially, psychologically or morally inappropriate.

Compulsive gambling is the phenomenon that sits most uncomfortably with this paradigm. 'Shopping addicts' don't behave like drug addicts in any sense *other than* devoting unusually large resources of time and energy to shopping.

²⁴⁵ See Koob, G. & Le Moal, M. (2000), Drug addiction, dysregulation of reward and allostasis. *Neuropsychopharmacology* 24:1-129.

²⁴⁶ Marijuana is the exception that proves the rule here. In many societies, both its financial and legal-risk costs are much lower than those of other illegal drugs. For this reason – and no other – many people are inclined to doubt that there is such a phenomenon as THC addiction.

They don't really experience withdrawal symptoms when they can't shop (regardless of how they might express their strong wishes to go shopping). They don't abandon their families or jobs as shopping consumes more and more of their lives. They don't dream obsessively of shopping or fantasize about ideal shopping scenarios, and they can drive past malls without scuttling important personal or business plans because of uncontrollable impulses to go in.²⁴⁷ This is what we would expect if 'shopping addiction' is just a metaphorical way of drawing attention to what a speaker regards as excessive indulgence in pleasure from shopping. But the point does not apply to gambling. Between 1% and 3% of the population of every country on which we have carefully gathered data both gambles heavily and frequently *and* exhibits the full set of behavioral properties with respect to gambling that are observed in those who are literally addicted to alcohol or heroin. The popular paradigm is therefore conflicted about the phenomenon. Many people are inclined to the view that there are true gambling addicts. Since this commits them to belief that people can be addicted to things besides drugs, those who hold this view are then typically inclined to drop their skepticism about literal television addiction and sex addiction and so on. Many other people prefer to hold a firmer conceptual line and deny that gambling addiction is literally real, notwithstanding the bizarrely stereotyped behavior of some excessive gamblers.

Scientists don't follow the popular practice of deciding whether or not to endorse literal gambling addiction as a pure question of *policy*. However, science *has* been influenced by the cultural paradigm, in at least one important way: the overwhelming bulk of careful investigation of addiction has been investigation of substance addiction. For that reason alone scientists have joined the general public in being undecided about literal gambling addiction. In fact, scientists have had a further motivation for this attitude. It was not, until the advent of contemporary neuroscience, obvious that there is *any* 'literal' addiction, in the sense of there being a firm and stable basis for dividing people into those who are genuine addicts and those who are not.²⁴⁸

Behavioral economists have been particularly wary of the concept, because it long seemed to rest on hypothesizing a special 'inner' property had by addicts but not by non-addicts, but on the basis of purely behavioral evidence.

Neuroscience has altered this situation. The I-RISA model literally describes particular relationships between particular people's brains and particular sources of reward for those people.

²⁴⁷ We are not claiming that absolutely *no one* behaves in these ways with respect to shopping (or sex or television or stamp collecting or anything else that most or some people enjoy). Addiction is a familiar cultural syndrome and so some people inevitably adopt it as a norm around which to structure their lives; they can, in principle, construct this norm around more or less anything. In addition, there are forms of psychological pathology that cause people to exhibit *some* of the core properties of addiction concerning an everyday activity, and these properties can be overwhelmingly vivid. Our claim is instead that there are vanishingly few cases of people who consistently act as if they're truly addicted to shopping, and there are whole cultures in which no one so acts. This strongly contrasts shopping with both drugs and gambling.

²⁴⁸ Of course any scientist would always have agreed that there are people whose lives are consumed and shattered by alcohol and cocaine, that many people struggle mightily and unsuccessfully in exerting effort not to smoke, and so on. But people's lives are consumed by all sorts of things, including politics and business. Few if any ever imagined that *that* should be diagnostic of addiction.

One can distinguish a nicotine or cocaine addict from one another, and from a non-addict, by examining their brains under fMRI or other scanning technologies as they are cued to anticipate different experiences. In principle at least, we could pick out addicts with enough of a history of use by examining the details of the connections between brain cells in their prefrontal cortices and in their amygdalas. Soon, we will very likely be able to diagnose addiction on the basis of genetic investigation. Researchers are starting to identify some of the genetic determinants of how reward systems develop, including the early response genes and postsynaptic density proteins involved in reward learning. Some of these proteins are altered in long-term ways by even single exposures to the classically addictive drugs. Exposure of recovered rat morphine addicts (but not normal rats) to drug-paired environments produces expression of a particular gene (*c-fos*) in several brain areas: medial prefrontal, ventrolateral orbital and cingulate cortex.²⁴⁹ There are similar findings for cocaine, amphetamine, nicotine, beer and palatable food²⁵⁰ (at least when the animals are in novel environments²⁵¹). There is evidence that all of this applies to some people with respect to gambling.²⁵² It probably also applies to some people with respect to some foods, but there is no evidence that it applies to anyone with respect to shopping, model railroad building, or sex.²⁵³

From recognition of gambling as a genuine target of addiction in some people, we will now argue for a stronger point. This is that AG emerges under the combination of the reward leaning / predictor valuation neuroeconomic model and the I-RISA neurochemical model as the template of 'basic' addiction, the form on which all other addictions are complications. This has been obscured by the cultural accident discussed above that scientific study of addiction began by focusing on drugs. However, there is now a neuroscience of AG. Furthermore, this neuroscience is rapidly forcing open the door to effective neuropharmacological interventions (i.e., drugs that act on the brain) for AG. This is just what we should expect if AG is the *basic* addiction.

²⁴⁹ Kelley, A. (2004). Memory and addiction: shared neural circuitry and molecular mechanisms. *Neuron* 44: 161-179.

²⁵⁰ Kelley, as immediately above.

²⁵¹ Wise, R. (2000). Addiction becomes a brain disease. *Neuron* 26: 27-33.

²⁵² Ibáñez, A., Blanco, C. Perez de Castro, I., Fernandez-Piqueras, J., & Sáiz-Ruiz, J. (2003). Genetics of pathological gambling. *Journal of Gambling Studies* 19: 11-22.

²⁵³ There is no shortage of popular, therapeutic, and non-neuroscientific social science discussion of, and practice oriented around, so-called 'sex addiction'. But there is no neuroscientific evidence for it, and, for reasons we have indicated, that is the only source of evidence we would regard as persuasive. Obviously, there are people who devote great time and energy to pursuit of sex. This is not evidence of addiction. Sex has a number of crucial disanalogies with targets of addiction. One of these is that fine manipulation of the reward contingency is not under an individual's own control. Admittedly, this objection does not apply to masturbation or devotion of attention to pornography. It is possible that there are masturbation and/or pornography addicts; but, again, we are aware of no neuroscientific evidence for this at present. Widespread belief in sex addiction is troubling. Neurochemical therapies for addiction are about to become common, as we will discuss below. There are grounds for concern that some people (especially minors) whose sexual behavior offends others' moral opinions will be inappropriately prescribed drugs that could damage their general motivational structures. This may be especially dangerous if, because they aren't really addicted to sex, they show little behavioral modification at normal doses for addictive substances and AG. We were motivated to add this note by concern that our endorsement of one form of non-drug addiction, AG, might be taken to imply endorsement of others that some parts of the therapeutic industry take to be similar to it.

Addictions will prove more difficult to treat the larger the number of different brain circuits they implicate; alcohol addiction, for example, is likely to prove a more stubborn treatment target than cocaine addiction because cocaine affects the reward system in a simpler way than alcohol. In this section and the next two we will explain why we are optimistic concerning the prospects for neuropharmacological management of AG.

We will present the case for this in two stages. First we will consider some theoretical reflections by neuroscientific experts on the nature of addiction that lead up to our interpretation. Then, in the next section, we will review the existing empirical neuroscience of AG, demonstrating that it has made remarkable progress very quickly. This will lead directly to a review of prospective neuropharmacological technologies.

The fact that neuroscience identifies distinctive properties that allow us to distinguish drug addicts, and then to associate AGs with them, still leaves us with puzzles about the development etiology (that is, the causal process in a particular person) and onset of addiction. The possibility of gambling addiction is bound to seem unlikely under any account of addiction that *begins* with drugs producing long-term changes in the brain as a consequence of the chemical properties of the drugs. As Wise²⁵⁴ argues, this includes not just traditional accounts that focus on chemical dependence and withdrawal, but also careless applications of current science that identify the onset of addictions with neuroadaptations (changes in synapses or connections between neurons). Wise emphasizes that rewarding properties of direct stimulants are learned extremely quickly; after as few as two or three electrical stimulations to the hypothalamus in their brains, rats will begin frantically pressing levers for more and ignoring food and water. Obviously, neuroadaptations cannot yet have occurred at the point where pursuit of the reward appears to be compulsive. “While it is clearly the case that brain changes associated with tolerance and dependence and brain changes associated with sensitization can develop under the right circumstances,” Wise points out, “what remains to be determined is the degree to which either of these changes is necessary for motivational habits to become compulsive. The rapid onset of compulsive self-stimulation [in rats and primates] would seem to preclude any of these drug-induced long-term neuroadaptations as a necessary condition for compulsive drug-seeking”. And it seems quite implausible that sensitization to gambling causes compulsive gambling. It appears much more likely that compulsive drug use or gambling causes sensitization.

This is why the I-RISA model of the neurochemistry must be complemented by the reward learning / predictor-valuation neuroeconomic model discussed in sections 3.1 and 3.2. The distinctive properties of addictive targets are not chemical properties *in the first place*. Rather, the crucial requirement on an addictive target is that it trips the reward circuit into what economists call hysteresis: a self-amplifying causal process driven by positive feedback. In the case of drugs, their chemical influence on Nacc dopamine levels *happens to be* the mechanism by which hysteresis is induced. If this were the *only* mechanism for such hysteresis, then there would be no harm in our identifying addictive properties with intrinsic chemical properties. But the phenomenon of AG refutes this conceptual move. The intrinsic chemical properties of drugs are, in this sense, *distractions* from the elementary structure of addiction. They are of course important, and we should expect that they will complicate – perhaps massively – development of neuropharmacological responses to addictive drugs. But sound method in both pure and applied science directs us to deal with basic cases of phenomena first and turn to complications later.

²⁵⁴ Wise, R. (2002). Brain reward circuitry: Insights from unsensed incentives. *Neuron* 36: 229-240. Quotations taken from pages 233 and 234.

These considerations suggest that AG should move to a position of central focus in addiction research. In their summary of the I-RISA model Goldstein and Volkow²⁵⁵ anticipate this point when they say that “a paradigm that simulates compulsive behavior (such as gambling when it is clearly no longer beneficial) might offer invaluable insight into the circuits underlying loss of control in addiction”.

In an addiction science that takes AG as the template instance, then, neurochemistry does not *supercede* neuropsychology and neuroeconomics, but complements them. Our attention is directed, in the first place, to understanding reward as a general process. Common sense is quite unfamiliar with this idea, since common sense identifies reward with either receipt of a tangible benefit (a nutritious meal, a monetary ticket to further consumption, a status-enhancing social acknowledgment, etc.), often including shots of hedonic delight. However, as Wise stresses, rewards as neuroscience requires us to understand them as “unsensed incentives”. The point of this phrase, in the context of reflections on targets of addiction, is that although people sense the *hedonic* consequences of drugs, they are not directly conscious of their rewarding properties. (This is most obvious in the case of nicotine, which is strongly addictive despite the fact that its episodic hedonic properties are not vivid to experienced smokers.) In the case of AG, we are apt to wonder what it is that the gambler is getting that keeps attracting her to gamble more. Is it thrills or money or relief from boredom or what? It now seems that this is the wrong question to ask. Once we understand reward in the scientific – neuroeconomic – sense, then there is enlightenment in saying: *the problem with gambling is that, for some people, it is rewarding in itself*.

This statement will appear trivializing rather than illuminating if the reader does not yet appreciate how different the neuroeconomic concept of reward is from folk conceptions. As Wise points out, in neuroscience the distinction between a reward and a predictor of a reward is subtle, perhaps nonexistent. Rewards are units of information that “by association establish otherwise neutral stimuli as things to be approached”. The reward system is a device for leading animals to approach certain things rather than others, and it encodes no distinction between what is of ultimate benefit and what predicts an experience the animal is conditioned to repeat. In the case of animals that must adapt to changing environments, the system has a bias for – takes as rewarding – any predictor of *novel* experiences that are not aversive. People, furthermore, are the ultimate novelty-seekers, our freakishly large brains signifying our status as the *only* animals in whom exaggerated drive for cognitive exploration is the basic species-specific adaptation. This emerges directly in the dopaminergic system’s design: dopamine is not released by familiar stimuli but by positive surprises. Furthermore, as Wise reminds us, since the *sensed* properties of the ‘reward’ in the folk sense (i.e., the food in the mouth, the orgasm, the drug high, the receipt of the winnings at the window) are not typically surprising, the system becomes unresponsive to them. So saying as we do that the dopaminergic system responds to predictors of reward is actually a bit misleading. What we’ve been calling ‘predictors of reward’ are (when positively surprising) the system’s rewards.

We think that this perspective goes a long way toward lifting the veil of mystery from AG. What gambling fundamentally consists in is paying for the possibility of a surprise. (This requires that odds in one’s favor not be *too* high, and it may be important to the reward that they’re negative. This doesn’t imply that the gambler doesn’t prefer winning to losing.

²⁵⁵ Goldstein, R., & Volkow, N. (2002). Drug addiction and its underlying neurobiological basis: neuroimaging evidence for the involvement of the prefrontal cortex. *American Journal of Psychiatry* 159: 1642-1652.

Note that nothing we have learned in psychology requires us to suppose that people's preferences are generally coherent, and much forces us to believe the opposite. All people who like gambling, and not just DGs, want two things at once that they can't have both of: to win every bet, and to participate in processes where there's a high risk of losing. Should this really surprise us? How many people want, in general, exciting lives but in which nothing seriously threatens them?) Thus the gambler is literally paying to engage his or her reward system in the most direct and straightforward possible way. In light of this insight, is it still surprising that when people discover that there are institutions where one can buy such direct manipulation of the dopaminergic system, hysteresis is induced in some of them? We suggest that the real surprise should be that *anyone* who gambles can *avoid* doing so pathologically. That is, if parts of us really are built to seek out surprise, and gambling produces opportunities for surprise with such convenience, it is perplexing that we don't all end up like the rats pressing levers which lead to stimulation of their hypothalami to the neglect of food and water.

The reader might still think we're making a lot of noise about the obvious. So for the precise sense in which puzzles arise for neuroeconomists *before* the insight, consider a sophisticated discussion by Ahmed published in 2004.²⁵⁶ Reviewing a neuroeconomic model of cocaine addiction developed by Redish, also published in 2004,²⁵⁷ Ahmed points out that Redish's model depends on the fact that cocaine directly causes dopamine release in Nacc, and so predicts strong reward regardless of any other contingencies. But then, Ahmed wonders, what about "addictions to ordinary rewards, such as fatty foods, which, unlike cocaine, produce a dopamine signal that can be accommodated"? We doubt that people are addicted to fatty foods; that many overindulge in them is no evidence at all for addiction. But in the clearest case of an ordinary reward to which people *do* become addicted, gambling, the answer to Ahmed's question is straightforward. Given what reward *means* in the mesolimbic system, and given what gambling *is* from this point of view, it too is guaranteed to predict strong reward come what may.

We hope it is now clear what we mean when we say that AG is the basic template of addiction. We mean more than just that it is the one stereotypical form in which there is no accompanying intrinsic chemical complication. We are alluding also to the fact, as explained immediately above, that gambling simply *is* direct stimulation of the mesolimbic reward system. We thus suspect that the syndrome of AG is relatively simple. We will now review preliminary evidence that we think bears this out. If our hunch is right, then AG may turn out to be the first major addiction we can pharmacologically manage.

3.5. The neuroscience of pathological gambling

In this section we will summarize results from the empirical neuroscientific research that has been done on PGs (most of whom, for reasons indicated in section 3.1, were probably AGs). This work comes from a variety of branches of neuroscience, each with a slightly different tradition of interpretation. We have pulled them together under *our* principles of interpretation, those of neuroeconomics as described in the preceding sections of this chapter. Thus the authors of the studies we cite might not always agree with our take on them, though there would of course be no disagreements about the factual results.

²⁵⁶ Ahmed, S. (2004).²⁵⁶ Addiction as compulsive reward prediction. *Science* 306: 1901-1902.

²⁵⁷ Redish, A. (2004),²⁵⁷ Addiction as a computational process gone awry, *Science* 306: 1944-1947

In light of the conclusions reached in the preceding sections of the report, the first question we must clearly ask about the empirical neuroscience of PG is: do we have clear evidence of abnormal subcortical and prefrontal dopamine levels (as we know are found in substance addicts) in PG subjects? The answer is 'yes'.²⁵⁸ In a 2004 review of neuroscientific and other biobehavioral studies of PGs, Goudriaan and colleagues²⁵⁹ mention some further supporting evidence implicating reward-system pathology in PGs. A sample of them had lower average baseline systolic blood pressure while gambling in the lab than controls. As measured by skin conductance, however, PGs were more aroused than controls by gambling cues. In general, however, physiological-behavioral studies of PGs until very recently measured effects that were too crude to discriminate among neural pathways or processes. Many behavioral studies are also compromised by offering no real stakes, or very small stakes, or absence of normal environmental cues. To the extent that AG is a disorder of the reward system, these are all serious methodological objections: small or absent stakes are unlikely to arouse a dopaminergic system that has protracted experience of gambling, while abnormal environments present the opposite problem: the reward system might be stimulated by learning novel association cues with gambling. This is why we have said a number of times that in the absence of research reported in the past two years, we would not even have been in a position to say whether there appeared to be any such phenomenon as addictive gambling.

The first neuroimaging study of PGs that showed much of strong relevance to the contemporary research setting was that of Stojanov and colleagues in 2003.²⁶⁰ They compared the prepulse inhibition (PPI) of PGs to PPI of controls. PPI is a measure of the reduction of brain activation to a stimulus S2 when a distracting stimulus S1 (a noise tone) is presented immediately before it. Researchers examine the acoustic startle reflex, which is an indicator of 'sensory motor gating' that inhibits processing of S2 so that S1 may be properly interpreted by the brain. Dopamine is the neurotransmitter mediating PPI. In Stojanov and colleagues's experiment, PGs showed reduced PPI relative to controls. This is evidence *either* of increased endogenous dopamine activity, or of higher serotonin or glutamate response. As indicated in section 3.3 above, all three of these possible explanations are consistent with what we see in the neurochemistry of substance addiction. The authors argue that the explanation most consistent with other standing data is that reduced PPI results from enhanced endogenous dopamine activity rather than reduced dopamine neurotransmission. Earlier work by Bergh and colleagues²⁶¹ had found that a sample of male PGs had reduced levels of dopamine in cerebrospinal fluid (CSF) relative to controls. Stojanov and colleagues appeal to this result in interpreting their data:

-
- ²⁵⁸ Blum, K., Braverman, E., Holder, J., Lubar, J.F., Monastra, V., Miller, D., Lubar, J.O., Chen, T., & Comings, D. (2000). Reward deficiency syndrome: a biogenetic model for the diagnosis and treatment of of impulsive, addictive and compulsive behaviors. *Journal of Psychoactive Drugs* 32: 1-112.).
- ²⁵⁹ Goudriaan, A., Oosterlaan, J., de Beurs, E., & Van den Brink, W. (2004). Pathological gambling: a comprehensive review of biobehavioral findings. *Neuroscience and Biobehavioral Reviews* 28: 123-141.
- ²⁶⁰ Stojanov, W., Karayanidis, F., Johnston, P., Bailey, A., Carr, V., & Schall, U. (2003). Disrupted sensory gating in pathological gambling. *Biological Psychiatry* 54: 474-484.
- ²⁶¹ Bergh, C., Eklund, T., Sodersten, P., & Nordin, C. (1997). Altered dopamine function in pathological gambling. *Psychological Medicine* 27: 473-475

“A more likely alternative [than reduced neurotransmission] is that pathologic gambling is associated with an increase in dopamine activity which, in turn, triggers a secondary increase in metabolism, thereby resulting in reduced CSF levels of dopamine”. It is known that dopamine D2 receptor agonists result in impaired sensory motor gating. Though there are identified genetic markers for this impairment, there is also animal-model evidence that it can be induced by stress during early development.

Non-technical synopsis: *In most people, if they are presented with two surprising stimuli close together, the brain blocks the dopamine system from paying too much attention to the second one so it doesn't interfere with completing processing of the first one. But too much dopamine activity can swamp this normal blocking. In the experiment, PGs didn't block the second stimuli as well as control subjects. The experimenters interpret this as evidence of higher-than-normal dopamine system activity in PGs, just as we find in drug addicts.*

In most of the other neuroimaging studies surveyed by Goudriaan and colleagues,²⁶² sample sizes were small, comorbidity factors were not assessed, and more PG subjects than controls had suffered brain traumas or taken psychotropic drugs. These are all severe compromising factors in efforts to compare PG with drug addictions.

Fortunately, neurochemical studies have been more informative. Five such studies have shown deficiencies in 5-HT (serotonin) levels. Two studies showed noradrenergic dysfunction. “[I]n two studies higher peripheral NE levels and blood levels of epinephrine were found during gambling and on gambling-concentrated days” Goudriaan and coauthors report. This is significant because Steketee²⁶³ cites evidence that epinephrine promotes dopamine flooding in VMPFC. Two studies found elevated dopamine levels in PG subjects, one by Shinohara and colleagues²⁶⁴ and the study by Bergh and colleagues mentioned above. There have also been a half-dozen failures to find differences in levels of other neurotransmitters. Thus it is suggestive that dopamine breaks through the somewhat murky data and reveals its familiar face.

Non-technical synopsis: *A number of neurochemical studies have found unusual transmitter levels in PGs that resemble those found in drug addicts. Furthermore, efforts so far to find unusual neurotransmitter profiles in PGs that don't resemble those of drug addicts have failed.*

In another study published in 2004 and hence too recent for inclusion in the survey by Goudriaan and colleagues, Meyer and colleagues²⁶⁵ found that PGs showed higher heart rate, norepinephrine, and dopamine levels than controls during a session of casino blackjack.

-
- ²⁶² Goudriaan, A., Oosterlaan, J., de Beurs, E., & Van den Brink, W. (2004). Pathological gambling: a comprehensive review of biobehavioral findings. *Neuroscience and Biobehavioral Reviews* 28: 123-141.
- ²⁶³ Steketee, J. (2002). Neurotransmitter systems of the medial prefrontal cortex: potential role in sensitization to psychostimulants. *Brain Research Reviews* 42: 203-228.
- ²⁶⁴ Shinohara, K., Yanigasawa, A., Kagota, Y., Gomi, A., Nemoto, K., Moriya, E., Furusawa, E., Furuya, K., & Teresawa, K. (1999). Physiological changes in Pachinko players: beta-endorphin, catecholamines, immune system substances and heart rate. *Applied Human Sciences* 18: 37-42
- ²⁶⁵ Meyer, G., Schwertfeger, J., Exton, M., Janssen, O., Knapp, W., Stadler, M., Schledowski, M., & Krüger, T. (2004). Neuroendocrine response to casino gambling in problem gamblers. *Psychoneuroendocrinology* 29: 1272-1280.

In Chapter 2 we reviewed the literature on performance of addicts and PGs in the Iowa Gambling Task (IGT) that is intended to measure cognitive bookkeeping as relevant to discounting. In one such experiment, Cavedini and colleagues²⁶⁶ scanned PGs and controls performing the IGT and other cognitive decision tasks with fMRI. Their PGs performed as well as normals on tasks that depended on generalized effective functioning of frontal lobes, but performed worse on the IGT. The authors suggest that this implies specific impairment of VMPFC. They go on to argue that “[t]hese data seem to suggest the existence of a link between PG, OCD and drug addiction, all having diminished ability to consider future consequences, which may be explained, at least in part, by abnormal functioning of the orbitofrontal cortex”. They note that “the literature suggests that deficit on decision making may reflect altered neuromodulation of the orbitofrontal cortex and interconnected limbic-striatal system by both the ascending 5-HT and mesocortical dopamine projections ... and that the functioning of the dopaminergic system, possibly mediating positive and negative reward, and the noradrenergic system, possibly mediating selective attention, is altered in pathological gambling”. The first suggestion is consistent with the I-RISA model’s mechanism for impaired response inhibition of the amygdala in addicts. The second, more general, suggestion of course points directly at the reward system, though we do not know why the authors disassociate the dopaminergic system from selective attention. This is perhaps merely a stylistic slip on their part.

Non-technical synopsis: *PGs performing the Iowa Gambling Task (discussed in Chapter 2) and other cognitive decision tasks were scanned with fMRI. They cognitively functioned as well as normal controls where thinking about the future wasn’t involved, but functioned worse than normal controls when the task required them to look ahead. The experimenters interpret these results, and the fMRI images of them, as suggesting that PGs have both of the main brain function abnormalities identified by the leading neurochemical model (I-RISA) of drug addiction.*

In a 2004 study Zack and Poulos²⁶⁷ used a novel methodology to investigate the extent to which stimulant drugs and gambling cues elicit common neural responses. They fed 30 mg. of D-amphetamine to four groups of subjects: (i) PGs, (ii) PGs comorbid for alcohol dependence, (iii) alcohol-dependent subjects who were not comorbid for PG, and (iv) healthy controls. They then assessed subjects’ subjective motivations to gamble and measured comparative reading rates under conditions where subjects were presented with cueing words from motivationally distinct semantic fields (gambling related, drinking related, positive affect, negative affect, neutral). Amphetamine, a drug whose stimulant properties are mediated by the dopaminergic system, increased self-reported motivation to gamble in PGs but not in controls. Furthermore, increased severity of PG predicted increased motivation under stimulant influence. By contrast, alcohol, as discussed earlier, exercises its drug effects by means of a different pathway, primarily involving GABA-A activation and glutamate inhibition, and working only indirectly on dopaminergic function. It is therefore interesting that amphetamines didn’t prime for drinking motivation in non-gambling alcohol abusers; this is some evidence that PG resembles stimulant drug action more than either resemble alcohol response, including in alcohol addicts. By the reading speed measure, stimulant drugs uniformly increased speeds across conditions in non-gamblers, while elevating attentiveness to gambling-related stimuli and reducing attentiveness to neutral stimuli in PGs.

²⁶⁶ Cavedini, P., Riboldi, G., Keller, R., D’Annunzi, A., & Bellodi, L. (2002). Frontal lobe dysfunction in pathological gambling patients. *Biological Psychiatry* 51: 334-341.

²⁶⁷ Zack, M., & Poulos, C. (2004). Amphetamine primes motivation to gamble and gambling-related semantic networks in problem gamblers. *Neuropsychopharmacology* 29: 195-207.

The authors relate this to a study by other researchers²⁶⁸ suggesting that reading speed differences are modulated by dopaminergic signals. These results are thus consistent with the suggestion that PG neurochemically resembles stimulus drug dependence, and does so to a greater degree than it resembles alcohol dependence.

Non-technical synopsis: *Experimenters fed amphetamines to a group of PGs, a group of PGs who were also alcoholics, a group of non-PG alcoholics, and a group of healthy controls. Then they gave all groups gambling-related, drinking-related, happy, sad and neutral cues. They asked subjects to report on their subjective motivations to gamble and to drink under each cue. Only PGs reported increased motivation to gamble when given the gambling cues under amphetamines, and the more severe their level of PG, the more their motivation to gamble was increased. By contrast, alcoholics didn't show increased motivation to drink under the influence of amphetamine. Subjects were also given reading tasks after the cues, and their changes in reading speed were measured. In all the non-PGs, including the alcoholics, reading speeds increased under amphetamines for all cues. However, in the PGs reading speeds declined under amphetamines for the gambling-related (and neutral) cues. This is all very interesting because it is evidence that amphetamines and PG have similar affects on attention mechanisms – i.e., the reward system – that neither shares with alcohol. Amphetamines are stimulant drugs that work directly on the dopamine system, whereas alcohol affects the dopamine system only indirectly.*

In related work, and using a complex experimental design, Potenza and colleagues²⁶⁹ investigated the hypothesis that PG brain responses to gambling cues will look more like cocaine addicts' responses to cocaine cues than like the responses of subjects with obsessive-compulsive disorder (OCD) to triggering stimuli. The motivation for this is the traditional interpretation of PG as a manifestation of OCD, as contrasted with the recent evidence that paints it as an addiction. The expression of the hypothesis in the experimental setting was that PGs would show elevated limbic system responses rather than elevated cortico-basal-ganglionic-thalamic responses compared to controls. The (all male) subjects were shown videotapes of sad, happy and casino-gambling scenarios. They were instructed to push buttons to signal onset of emotional and motivational episodes, and were asked to rate peaks and intensities of these episodes. The groups did not differ significantly in subjective responses to sad and happy scenarios. However, all PGs, but only a minority of non-PGs, reported gambling urges while viewing gambling scenarios. Before subjective response in the gambling scenarios, PGs showed decreased activity, and controls increased activity, in OFC, basal ganglia and thalamus. Sad and happy scenarios elicited no significant such differences. At time of subjective response, PGs showed increased activity relative to controls in OFC for gambling scenarios, and in ventral striatum and OFC for happy scenarios. During the final period of viewing, after presentation of the scenarios, PGs had decreased activity in some regions for sad scenarios relative to controls, and decreased activity in ventral anterior cingulate for gambling scenarios. Other studies have associated this with positively valenced mood. So, on the one hand, anticipatory cue responses in PGs differed from those of OCD subjects (who go the opposite way); but the later decreased activity in anterior cingulate distinguished PGs from cocaine abusers. With respect to the role of the reward system, the first result associates PG with addiction.

²⁶⁸ Kischka, U., Kammer, T., Maier, S., Weisbrod, M. Thimm, M., & Spitzer, M. (1996). Dopamine modulation of semantic network activation. *Neuropsychologia* 34: 1107-1113.

²⁶⁹ Potenza, M., Steinberg, M., Skudlarski, P., Fulbright, R., Lacadie, C., Wilber, M., Rounsaville, B., Gore, J., & Wexler, B. (2003). Gambling urges in pathological gambling. A functional magnetic resonance imaging study. *Archives of General Psychiatry* 60: 828-836.

Since anterior cingulate has been implicated in decision-making relevant to control of impulses, the second result has PGs conforming *better* to the I-RISA model than cocaine addicts.

Non-technical synopsis: *PG has traditionally been thought to be a variety of obsessive-compulsive disorder (OCD). The experimenters designed tests to find out whether a group of PGs would respond more like OCDs or more like cocaine addicts. PGs, OCD subjects and normal controls were shown happy, sad, and gambling-related scenarios. They were asked to push buttons when they felt emotional and motivational state changes, and were asked to rate the peaks and average intensities of each of these episodes. During anticipation phases before gambling scenarios, but not happy or sad ones, PGs showed the opposite response to OCD subjects and controls. This is the same response that cocaine addicts show in anticipation of cocaine-related stimuli. PGs also showed decreased activity after gambling-related scenarios in a brain region where cocaine addicts don't show decreased activity after cocaine-related stimuli. But here, it's the PGs, and not the cocaine addicts, who are behaving in a way that is explained by the standard (I-RISA) model of drug addiction.*

Of interest following on from the first result is an fMRI examination published in 2005 by Crockford and colleagues²⁷⁰ of male PGs and controls while they watched, in alternation, video images of gambling scenarios and videos of nature scenes. Their study was motivated by the expectation that if mesolimbic processing of motivation associated with reward expectancy in PGs were activating OFC, this should be accompanied by elevated activity in dorsolateral prefrontal cortex (DLPFC) so as to recruit working memory to track the expected gambling-relevant information. In the experiment, PGs indeed showed elevated DLPFC activity relative to controls while watching the gambling videos, but not while engaged with the nature videos. In addition, PGs showed elevated activity relative to controls in right parahippocampal region and left occipital cortex during the gambling presentations. These elevated activations correlate with subjective reports of cravings to gamble by the PG subjects. These results are interesting in light of a 2004 finding by Dorris and Glimcher²⁷¹ that posterior parietal cortex (PPC) encodes transformations of representations into generation of movement, since DLPFC is reciprocally linked with PPC. Subjective cravings may in part be the experience of preparation for action.

Non-technical synopsis: *The standard (I-RISA) model of drug addiction predicts that when the reward system of the addict is presented with cues for its drug, the part of the brain responsible for working memory should show more activation than in non-addicts. PGs and normal controls were shown gambling-related and neutral videos. The working memory area in the PGs indeed showed higher activation than the same area in the controls for the gambling-related but not the neutral stimuli. PGs also showed activation in a part of the brain that exchanges signals with the part responsible for initiation of action, and this activity was correlated with their subjective reports of gambling cravings. So cravings might in part be the AG's experience of preparation for action to bring about a gambling episode.*

²⁷⁰ Crockford, D., Goodyear, B., Edwards, J., Quickfall, J., & el-Guebaly, N. (2005). Cue-induced brain activity in pathological gamblers. *Biological Psychiatry* forthcoming.

²⁷¹ Dorris, M., & Glimcher, P. (2004). Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron* 44: 365-378.

In a study shedding further light on impairment of response inhibition, Potenza and colleagues²⁷² scanned PGs and controls under fMRI while they performed a 'Stroop task' (see above). PGs differed from controls only in showing decreased activity in VMPFC – once again suggestive of reduced capacity for impulse control and conforming to the prediction of the I-RISA model for addicts.

Non-technical synopsis: *While performing another cognitive task under fMRI, PGs showed decreased activation in a part of the brain predicted by the I-RISA model of drug addiction to show decreased activation in addicts.*

A related result, also published in 2005 was found by Reuter and colleagues²⁷³ when they scanned PGs and controls under fMRI while they engaged in a guessing game for prizes. PGs showed decreased activation in ventral striatum on wins, just as drug addicts do. They also showed decreased VMPFC activity, complementing some of the results described above (those of Cavedini and colleagues, and Potenza and colleagues) and predictive of impaired response inhibition in accordance with the I-RISA model. The authors interpret their data in a way consistent with the I-RISA model as applied to PGs, but going beyond it in also offering a hypothesis about the causal background to development of addiction: "Drug self-administration experiments have suggested that organisms try to maintain a homeostatic baseline level of dopamine in the ventral striatum, which in normal volunteers can be maintained by weak reinforcers found in everyday life. In contrast, in PGs (who have reduced striatal activation), natural reinforcers are not strong enough for dopamine to reach and maintain this homeostatic baseline level. At this stage organisms seek additional, stronger reinforcers, such as drugs of addiction or gambling, to compensate for the lack of activation". This hypothesis rather outruns available data in our opinion, though it has some very general basis in the general neuroscience literature on addiction.²⁷⁴

Non-technical synopsis: *Experimenters scanned groups of PGs and controls under fMRI while they played a guessing game for prizes. Once again, PGs showed brain activation patterns as predicted for addicts by the I-RISA model. The experimenters interpret this in terms of a theory according to which addicts are people whose dopamine levels can't be maintained at the naturally appropriate level by normal rewards. In our view, this hypothesis reaches beyond what current data support.*

In section 3.3 we described evidence for a more modest speculation on the developmental basis of addiction, the 2003 work of Chambers, Taylor and Potenza²⁷⁵ on correlations between age-related vulnerability to addiction and the stages of development of the human brain.

²⁷² Potenza, M., Leung, H.-C., Blumberg, H., Peterson, B., Fulbright, R., Lacadie, C., Skudlarski, P., & Gore, J. (2003). An fMRI Stroop task study of ventromedial prefrontal cortical function in pathological gamblers. *American Journal of Psychiatry* 160: 1990-1994.

²⁷³ Reuter, J., Raedler, T., Rose, M., Hand, I., Gläscher, J., & Büchel, C. (2005). Pathological gambling is linked to reduced activation of the mesolimbic reward system. *Nature Neuroscience* 8: 147-148.

²⁷⁴ Koob, G. & Le Moal, M. (2000). Drug addiction, dysregulation of reward and allostasis. *Neuropsychopharmacology* 24:1-129.

²⁷⁵ Chambers, R.A., Taylor, J., & Potenza, M. (2003). Developmental neurocircuitry of motivation in adolescence: A critical period of addiction vulnerability. *American Journal of Psychiatry* 160: 1041-1052.

As we saw, the adolescent brain, which is selected for relative emphasis on exploration, has underdeveloped cortical response inhibition machinery compared to the typical adult brain. In follow-up work Chambers and Potenza²⁷⁶ review the relationship between this correlation and data on PG. As we've noted PG rates are higher among adolescents than among adults, despite less convenient access to gambling venues and less disposable income in the former. Furthermore, PG in adolescence is a strong predictor of PG in adulthood.²⁷⁷ Thus it is almost certainly significant that, as the authors say, "[I]n adolescents as compared with adults promotional mechanisms [i.e., the dopaminergic system] are relatively hyperactive and inhibitory mechanisms [i.e., OFC and 5-HT circuits] relatively hypoactive". Unfortunately, no direct fMRI studies of adolescent PG subjects have yet been reported.

To the extent that we have grounds for looking for the developmental etiology of AG in pre-adulthood (including in pre-natal genetic background), we may be interested in 'twin data' on PG. So-called 'twin studies', in which fraternal twins who grew up in the same household are compared with fraternal twins raised separately with respect to a dependent variable, have long been employed in developmental psychology as a way of exploring the relative causal strengths of genetic and environmental influences. Such a study has been conducted for PG by Eisen and colleagues²⁷⁸ on 3359 twin pairs. The authors report that inherited factors explain between 35% and 54% of reports of five DSM symptoms of PG that could be estimated statistically. Familial factors explain 56% of the report of three or more symptoms, and 62% of PG diagnoses. We recommend extreme caution with respect to these conclusions, since the regression methods involved in generating the statistical estimations are extremely sensitive to small differences in probabilities of autocorrelations, about which little, in the case of estimators for PG, is known.

Non-technical synopsis: *Researchers studied over 3,000 identical twins of PGs, some of whom had been raised apart from the PG and some of whom had been raised with the PG, to determine whether the twin had also become a PG. This is a standard method used to try to assess the relative influence of genetic and environmental factors on an outcome. They found that if one does not suppose there is an inherited causal factor for PG, then prevalence of diagnosed PG in the twins would be predicted to be between one third and two thirds lower than what the researchers found. The authors are cautious in interpreting this result, as it is highly sensitive to facts about the distribution of the DSM diagnostic properties for PG about which little is known. We urge extreme caution.*

Of much more interest, in our view, is preliminary work that directly examines the genomes of PGs. Pioneering work has been carried out by Ibáñez and colleagues.²⁷⁹ There are, in general, two approaches to trying to isolate genes for specific properties or pathologies.

-
- ²⁷⁶ Chambers, R.A., & Potenza, M. (2003). Neurodevelopment, impulsivity and adolescent gambling. *Journal of Gambling Studies* 19: 53-84.
- ²⁷⁷ Winters, K., Stinchfield, R., & Fulkerson, J. (1990). Adolescent gambling behavior in Minnesota: benchmark. *Report to the Department of Human Services Mental Health Division*. Duluth: Center for Addiction Studies, University of Minnesota.
- ²⁷⁸ Eisen, S., Lin, N., Lyons, M., Scherrer, J., Griffith, K., True, W., Goldberg, J., & Tsuang, M. (1998). Familial influences on gambling behavior: an analysis of 3359 twin pairs. *Addiction* 93: 1375-1384.
- ²⁷⁹ Ibáñez, A., Blanco, C., Perez de Castro, I., Fernandez-Piqueras, J., & Sáiz-Ruiz, J. (2003). Genetics of pathological gambling. *Journal of Gambling Studies* 19: 11-22.

Linkage studies, which are typically the most powerful, look for marker alleles within families that distinguish affected from unaffected family members. However, very often samples of related people large enough relative to expression rates of the pathology cannot be found. Thus genetic detective work more often relies on association studies, which look for statistically significant correlations between polymorphic alleles (sites on the DNA molecule where different possible expressions have been identified and statistically estimated) and expression of the pathology in an affected group relative to a control group from the same population. Ibáñez and colleagues indicate that no linkage studies of PG had been reported as of their writing. However, seven association studies had been done, of which all but one had been conducted by the Ibáñez research group. A very helpful property of their work is that they motivate their hypotheses by efforts to provide separating evidence between conceptions of PG as an obsessive-compulsive disorder mainly implicating the noradrenergic system and the addiction (I-RISA) model that directs attention to the dopaminergic pathways. “Consequently,” they say, “genes relevant to the function of serotonergic, dopaminergic and noradrenergic systems could be considered as candidate genes in PG”. They found a DNA sequence (a promoter polymorphism) for expression of a protein known as MOA-A that has been associated with control of neurotransmitters found in the reward system to be significantly increased in male PGs as compared to controls. “Interestingly,” they comment, “although serotonin is a preferred substrate for MOA-A, MOA-A is expressed in the brain mainly in dopaminergic neurons, raising the question of whether these allele variants are more likely to result in changes in serotonergic or dopaminergic transmission.” The group also studied a polymorphism in the D2 dopamine receptor gene “and found significant differences in allele distribution in PGs with as compared to those without co-morbid psychiatric disorders”. In female but not male PGs, they discovered²⁸⁰ the DRD4 7-repeat allele, coding for less efficient receptors, to be significantly more frequent than in controls. Finally, they report that in the study carried out by a different team,²⁸¹ researchers “found a statistically significant association between the Taq-A1 allele of the D2 dopamine receptor gene in pathological gamblers compared to controls. The Taq-A1 allele has also been found to be associated with other impulsive-addictive-compulsive behaviors, leading some researchers to propose a *Reward Deficiency Syndrome* as an underlying genetic foundation for these disorders.”²⁸² Thus the prospect looms that genetics could explain the dynamics described by the I-RISA model, and some or perhaps even most AG could turn out to have a genetic etiology. This would of course have massive implications for prevention possibilities, as we stand at the dawn of the age of genetic screening and therapies. It should be noted that more than one genetic syndrome will likely turn out to be implicated, since genetic correlations with PG found to date have tended to be sex-specific.

-
- ²⁸⁰ Perez de Castro, I., Ibáñez, A., Torres, P., Sáiz-Ruiz, J., & Fernandez-Piqueras, J. (1997). Genetic association study between pathological gambling and a functional DNA polymorphism at the D4 receptor. *Pharmacogenetics* 7: 345-348
- ²⁸¹ Comings, D., Rosenthal, R., & Lesieur, H. (1996). A study of the dopamine D2 receptor gene in pathological gambling. *Pharmacogenetics* 6: 223-234.
- ²⁸² The “some researchers” to whom they refer are: Blum, K., Sheridan, P., Wood, R., Braverman, E., Chen, T., & Comings, D. (1995). Dopamine D2 receptor gene variants: Associative and linkage studies in impulsive-addictive-compulsive behavior. *Pharmacogenetics* 5:121-141.

This might, indeed, explain why PG prevalence among males is reliably higher, in all countries where statistical evidence has been carefully collected, despite the fact that differences in testosterone levels do not appear to be related to dispositions to PG.²⁸³

Non-technical synopsis: *Researchers have investigated PG subjects in search of a possible genetic basis for PG. Seven studies have compared the genomes of PG subjects to those of non-PG subjects drawn from the same populations. They have explicitly focused this work on genes which, if discovered to be correlated with PG, would help us decide between the hypothesis that PG is a form of obsessive-compulsive disorder and the hypothesis that many, most or all PGs are addicts. These studies have found a gene that has been associated with variance in both serotonin and dopamine to be correlated with PG in males, correlations of two other genes related to dopamine production with PG in both sexes, and correlation between a gene for inefficient neurotransmission and PG in women. These results all favour the addiction model. Thus there might be a genetic basis for some, most or all AG. Since some genetic factors so far related to PG are sex-specific, this might partly explain why there are more male than female PGs, even though differences in testosterone levels do not appear to be related to the probability of a person's being a PG.*

The evidence reviewed in this chapter, though based on a still small number of studies, speaks unanimously in support of the claim that PGs have the suite of neurochemical properties in common with drug addicts that the I-RISA model of addiction holds to be crucial. Though more research is clearly needed (and is clearly forthcoming), it is unusual for a initial data for a new hypothesis to come in so univocally on the positive side. It seems very likely that there is such a thing as gambling addiction in the neurochemical sense of I-RISA, and that many or most PGs are in fact AGs.

3.6. Neuropharmacology of pathological gambling

On the basis of the evidence reviewed to this point, we regard the I-RISA model as the best confirmed account of the neurochemistry of AG, with the reward learning / predictor-valuation neuroeconomic model of reward providing the theoretical generalization and explanation of I-RISA. This tentative conclusion both has implications for, and is given further support by existing trials of, pharmacological treatments of PG subjects. At present there are no medications for PG approved by the US Food and Drug Administration, which can be expected to set the precedent for the rest of the world because the majority of relevant research is being conducted in America. On the drug therapies that look most promising at this time, there are as yet no controlled clinical trials to report, but only pilot studies. Were it not for the fact that we regard the combined I-RISA / neuroeconomic account of AG as well confirmed, our basis for optimism about chemical agents that have been successful in pilot studies would have little basis. Fortunately, however, the results of the pilot studies we will describe are in almost perfect accord with the explanatory models we have reviewed, which both strengthens the case for the models and improves the significance of the small pilot studies.

We begin by summarizing an early (2003) review prepared by Grant, Kim and Potenza.²⁸⁴

²⁸³ Blanco, C., Ibáñez, A., Blanco-Jerez, C.-R., Baca-Garcia, E., & Sáiz-Ruiz, J. (2001). Plasma testosterone and pathological gambling. *Psychiatry Research* 105: 117-121.

²⁸⁴ Grant, J., Kim, S., & Potenza, M. (2003). Advances in the pharmacological treatment of problem gambling. *Journal of Gambling Studies* 19: 85-109.

They consider classes of pharmacological interventions in roughly the chronological order in which they have been tested in PGs. There are four such classes: (1) mood stabilizers, (2) serotonergic agents, (3) opioid receptor antagonists,²⁸⁵ and (4) dopaminergic agents.

(1) Mood stabilizers (e.g., lithium, carbamazepine) have been sporadically used to treat PGs since the 1970s. Existing studies are too small and incompletely described to tell us whether this is helpful for PGs whose gambling problems aren't comorbid with a specific mood disorder.

(2) There is some evidence that serotonin reuptake inhibitors, such as clomipramine, reduce PG symptoms, particularly in males. This is intuitive in the light of the I-RISA model, since one of the two main neurochemical mechanisms that it posits as underlying PG is 5-HT deficit that expresses impaired cortical response inhibition of the amygdala. Furthermore, 5-HT agonists have been found to significantly modulate dopamine release in Nacc.²⁸⁶ If SRIs are effective against PG, then the newer generation of *selective* SRIs (SSRIs) are likely to be more so. Several small single-blind and double-blind trials of the SSRI fluvoxamine showed it as effective in controlling PG behavior, at least in the short term. Fluvoxamine has specifically been demonstrated to reduce extracellular dopamine levels in Nacc.²⁸⁷ Interestingly, in light of the role of serotonin in the I-RISA model, these effects seem to be independent of symptoms of depression, and doses for efficacy are higher than standard doses for treating depression. This is supported by a 1998 study conducted by Hollander and colleagues,²⁸⁸ though this work's interpretation is obscured by the fact that the researchers did not screen for depression comorbidity. It must be noted, however, that some comparison studies of SSRIs and placebos have showed significant short-term response in PGs to placebos. It is thus problematic that most cited studies of SSRI trials with PGs (e.g., in 2002 by Zimmerman and colleagues²⁸⁹) have not been placebo-controlled.

Non-technical synopsis: *Drugs for modulating the neurotransmitter serotonin have shown some promise in treating PG, especially in men. This makes sense if the I-RISA model correctly describes AG. However, we must be a bit cautious about these studies because many were not placebo-controlled, and PGs have been found to respond relatively strongly, in the short term, to placebos.*

(3) Interesting results have been obtained with opioid receptor antagonists, in particular, with naltrexone.²⁹⁰

-
- ²⁸⁵ An antagonist for a given neurotransmitter is a drug that blocks receptors for the neurotransmitter in question.
- ²⁸⁶ Boulenguez, P., Rawlins, J., Chauveau, J., Joseph, M., Mitchell, S., & Gray, J. (1996). Modulation of dopamine release in the nucleus accumbens by 5-HT_{1B} agonists: involvement of the hippocampo-accumbens pathway. *Neuropharmacology* 35: 1521-1529.
- ²⁸⁷ Ichikawa, J., & Meltzer, H. (1995). Effect of antidepressants on striatal and accumbens extracellular dopamine levels. *European Journal of Pharmacology* 281: 255-261.
- ²⁸⁸ Hollander, E., DeCarla, C., Mari, E., Wong, C., Mosovich, S., Grossman, R., & Begaz, T. (1998). Short-term single-blind fluvoxamine treatment of pathological gambling. *American Journal of Psychiatry* 155: 1781-1783.
- ²⁸⁹ Zimmerman, M., Breen, R., & Posternak, M. (2002). An open-label study of citalopram in the treatment of pathological gambling. *Journal of Clinical Psychiatry* 63: 44-48.
- ²⁹⁰ Kim, S., & Grant, J. (2001). An open naltrexone treatment study of pathological gambling disorder. *International Clinical Psychopharmacology* 16: 285-289.; Kim, S., Grant, J., Adson,

This antagonizes the mu-opioid system, activity of which is implicated, as we have seen by the I-RISA model, according to which diminished GABA in the mesolimbic system contributes to dopamine concentration in Nacc. Naltrexone directly modulates GABA input to dopamine neurons in the mesolimbic pathway. A preponderance of uses of this treatment on PGs have involved subjects with comorbid alcohol dependency. This is important, since it is known that alcohol addiction works its dominant effects through the GABAergic system. Thus until the first recent trial of naltrexone (on which, see more below) that compared its action in non-alcohol-dependent PGs with control populations,²⁹¹ there wasn't yet strong reason to regard it as a supported approach to PG *per se*. Note that there is a small window between the minimum dosage indicated as effective and dosages that impair liver function. Despite these grounds for caution on viewing opioid receptors as mainstream therapy for PG, the evidence of their efficacy supports the I-RISA model of addiction, and its specific application to PG.

Non-technical synopsis: *Another class of drugs that has been used with success to treat PGs in a small group of studies works on another neurotransmitter, GABA. This is a main neurotransmitter involved in alcohol and opiate addiction. This class of drugs might turn out to be especially useful for patients who are addicted to both gambling and alcohol. The window between minimum doses that have been found to be effective and doses that risk liver damage is small. Effectiveness of these drugs against PG is further evidence that the I-RISA model of addiction applies to gambling.*

(4) So-called 'atypical antipsychotics' (e.g., risperidone, olanzapine, quetiapine and ziprasidone), used for treating schizophrenia and other psychotic disorders, antagonize dopamine D2 receptors along with serotonin 5-HT_{2A} and / or 5-HT_{2C} receptors. Limited clinical results have shown that when these agents are used to treat PGs also diagnosed as schizophrenic, gambling behavior moderates significantly.²⁹² Grant and coauthors call for further studies of the efficacy of atypical antipsychotics in treating *non-schizophrenic* PGs. In light of the evidence reviewed in section 3.5, suggesting that PG is a direct disorder of the dopaminergic system, with subsequent implications for PFC 5-HT circuits, this call was clearly in order. As we will see in a moment, it is being answered.

Non-technical synopsis: *In light of the research reviewed in section 3.5 above, another class of drugs that clearly merit attention where PG is concerned are those that modulate dopamine. These are presently used to control schizophrenia.*

As Grant and colleagues indicate, just two years ago there was much more information available on how to pharmacologically treat PGs by reference to some of their typical comorbidities than to PG *per se*.²⁹³ If a patient is comorbid for depression, there are clearly grounds for treatment with an SSRI. In a PG with comorbid alcohol dependence, naltrexone therapy is an option favored by some evidence.

D., & Rummel, R. (2001). Double-blind naltrexone and placebo comparison study in the treatment of pathological gambling. *Biological Psychiatry* 49: 914-921.

²⁹¹ Dannon, N., Lowengrub, K., Musin, E., Gonopolski, Y., & Kotler, M. (2005). Sustained-release bupropion versus naltrexone in the treatment of pathological gambling: A preliminary blind-rater study. *Journal of Clinical Psychopharmacology* 25: 593-596.

²⁹² Chambers, R.A., & Potenza, M. (2001). Schizophrenia and pathological gambling. *American Journal of Psychiatry* 158: 497-498.

²⁹³ A similar point is made and documented by Dell'Osso, B., Allen, A., & Hollander, E. (2005). Comorbidity issues in the pharmacological treatment of pathological gambling: a critical review. *Clinical Practice and Epidemiology in Mental Health* 1: 21.

However, genuine pharmacological management of PG would require our knowing what generally works against PG symptoms in the first place.

A more recent survey of the neuropharmacology of PG is the 2005 account by Hollander and colleagues.²⁹⁴ They report two SSRI studies not mentioned in the 2003 review by Grant, Kim and Potenza. First, in 2003 Grant and colleagues²⁹⁵ tested paroxetine on PGs with comorbid axis-I disorders screened out. They found no evidence of efficacy as against placebos. In a 2002 study Pallanti and colleagues²⁹⁶ found nefazodone, which inhibits both serotonergic and adrenergic reuptake, to be efficacious in a small non-placebo-controlled trial.

Two recent studies have tested dopamine reuptake inhibitors (atypical antipsychotics) on PG subjects. First, Black²⁹⁷ treated 10 PGs with up to 300 mg/day of bupropion for 8 weeks. There was significant measured improvement for all 10 subjects. Earlier this month (from this writing), Dannon and colleagues²⁹⁸ reported on a comparative trial of bupropion and nexaltone (opioid receptor antagonist) with PGs. Each group was treated for 12 weeks. 'Full response' was defined as absence of gambling for 2 weeks following treatment, combined with improved score on the Clinical Global Impression-Improvement Scale. 'Partial response' was defined as a decrease in the frequency of gambling and a decrease in the stakes wagered. 75% of the bupropion-treated group and 76% of the nexaltone-treated group showed full response, and 100% of each sample showed at least partial response. Needless to say, these are encouraging results, the absence of placebo control notwithstanding.

Non-technical synopsis: *Two studies of the efficacy of drugs that modulate dopamine have now been reported. Both show 100% efficacy on a weak measure, while one drops to 75% efficacy on a stronger measure. The second trial tested the dopamine drug against a drug that acts on GABA. It showed identical efficacy.*

An intriguing and highly suggestive clinical development occurred over the past year when Dodd and colleagues²⁹⁹ encountered 11 Parkinson's disease sufferers without histories of even mildly disordered gambling who developed severe PG after stabilizing on courses of a dopamine agonist (pramipexole in 9 of the cases and ropinirole in the other 3 cases). In all cases, PG symptoms resolved rapidly after the medication was discontinued.

-
- ²⁹⁴ Hollander, E., Sood, E., Pallanti, S., Baldini-Rossi, N., & Baker, B. (2005). Pharmacological treatments of pathological gambling. *Journal of Gambling Studies* 21: 101-110.
- ²⁹⁵ Grant, J., Kim, S., Potenza, M., & Blanco, C. (2003). Paroxetine treatment of pathological gambling: a multi-centre randomized controlled trial. *International Clinical Psychopharmacology* 18: 243-249.
- ²⁹⁶ Pallanti, S., Baldini-Rossi, N., Sood, E., & Hollander, E. (2002). Nefazodone treatment of pathological gambling: A prospective open label controlled trial. *Journal of Clinical Psychiatry* 63: 1034-1039.
- ²⁹⁷ Black, D. (2004). An open-label trial of bupropion in the treatment of pathological gambling. *Journal of Clinical Psychopharmacology* 24: 108-110.
- ²⁹⁸ Dannon, N., Lowengrub, K., Musin, E., Gonopolski, Y., & Kotler, M. (2005). Sustained-release bupropion versus naltrexone in the treatment of pathological gambling: A preliminary blind-rater study. *Journal of Clinical Psychopharmacology* 25: 593-596.
- ²⁹⁹ Dodd, M.L., Klos, K., Bower, J., Geda, Y., Josephs, K., & Ahlskog, E. (2005). Pathological gambling caused by drugs used to treat Parkinson disease. *Archives of Neurology* 62: 1-5.

On the basis of this discovery, the researchers performed a literature search that revealed six published reports collectively indicating 17 Parkinson's patients who developed PG after beginning courses of dopamine agonists. Most of these patients had also been taking a dopamine stimulant, levodopa. However, none of Dodd and colleagues's sample developed PG while on levodopa alone, and 3 of the sample were not taking levodopa. The literature search sample implicates all of the commonly prescribed dopamine agonists, but pramipexole disproportionately (82% in Dodd and colleagues's sample and 59% in the search set). "Pramipexole," they comment, "is a unique drug that is highly selective for the dopamine D3 receptor, with an affinity at least two orders of magnitude greater than for other receptors". Pramipexole plus pergolide and ropinirole, which are also selective (though less strongly so) for D3 account for 93% of the 28 induced PG cases. Dodd and colleagues note that "this makes intuitive sense in view of the localization of D3 receptors to limbic areas of the brain".

Indeed there is intuitive resonance here, *especially* to the extent that we believe that (1) some or all PG is an addictive disorder as characterized by the I-RISA model *and* (2) that some or all AG mimics stimulant drugs as a syndrome that promotes its effects by hijacking of reward earning / predictor valuation through stimulation of dopamine elevation in Nacc. We have surveyed evidence for both (1) and (2). In addition, we find D3 receptors specifically implicated in some other work on addiction as a mesolimbic reward-pathway disorder. Pilla and colleagues³⁰⁰ found that targeting mesolimbic D3 receptors with a dopamine agonist "greatly" diminished cocaine-seeking behavior. Hypothesis (2) of course predicts that treatments against cocaine-seeking are suggestive models for AG therapies. Everitt, Dickinson and Robbins³⁰¹ note, a propos of this study, that "it remains unclear whether this effect is mediated by blockade of D3 receptors within the Nacc, the amygdala or autoreceptors on the mesolimbic [dopamine] neurons themselves." However, they add that "[o]ne advantage of drugs acting at the D3 receptor is that its distribution in the brain is restricted to these few sites ... and so unwanted side effects of such drugs are minimal." In addition, Chiamulera,³⁰² developing an I-RISA -style model of nicotine addiction as a reward-system disorder, reports in 2004 that "[d]opamine D3 receptor antagonist SB-277011 inhibited reinforcement of responding in a model of relapse to nicotine where responding was induced by the sequential presentation of CX/CS and then of a single nicotine injection." In support of this he cites Andreoli and colleagues.³⁰³

Non-technical synopsis: *The drugs that have turned some Parkinson's disease patients into PGs disproportionately acted on a very particular little group of neurons, the so-called 'dopamine D3 receptors' in one part of the brain (the part that is the prime target of addictions according to all models of current interest). And now, looking back over the literature, we notice D3 popping up regularly at the scene of the crime (addiction).*

³⁰⁰ Pilla, M., Perachon, S., Sautel, F., Garrido, F., Mann, A., Wemuth, C., Schwartz, J., Everitt, B., & Sokoloff, P. (1999). Selective inhibition of cocaine-seeking behaviour by a partial dopamine D3 receptor agonist. *Nature* 400: 371-375.

³⁰¹ Everitt, B., Dickinson, A., & Robbins, T. (2001). The neuropsychological basis of addictive behavior. *Brain Research Reviews* 36: 129-138.

³⁰² Chiamulera, C. (2004). Cue reactivity in nicotine and tobacco dependence: a "multiple-action" model of nicotine as a primary reinforcement and as an enhancer of the effects of smoking-associated stimuli. *Brain Research Reviews* 48: 74-97.

³⁰³ Andreoli, M., Tessari, M., Pilla, M., Valerio, E., Hagen, J., & Heidbreder, C. (2003). Selective antagonism at D3 receptors prevents nicotine-triggered relapse to nicotine-seeking behavior. *Neuropharmacology* 28: 1272-1280.

This is intriguing and encouraging, because the strong localization of these receptors offers hope for an anti-PG drug that might have minimal side-effects on general motivation, concentration and motor control.

A major challenge that will arise for efforts to pharmacologically manage AG, if indeed it is wholly or mainly a reward-system pathology, is identified by Baunez and colleagues.³⁰⁴ Because drugs of abuse and 'natural' reinforcers – of which, we argued in section 3.4, AG is the template – exert their effects via the same neural pathways and processes, "in drug addicts, one of the main unsolved challenges is eradicating the pathological motivation for the drug without impairing the general affective state of subjects". This is an especially troubling concern when our treatment target is itself a natural reward. Will it be possible to manage AG pharmacologically without inducing general motivational apathy in patients? In this context, it would be greatly encouraging if it turned out that anti-AG therapy could focus on a sparsely distributed receptor such as D3.

All of this is still merely suggestive. As Stacchi³⁰⁵ reminds us, only about 1.5% of Parkinson's patients who take pramipexole and 0.3% who take other D3 agonists develop PG. So we do not have evidence of a simple switch which, turned too far in one direction, produces Parkinson's disease, and turned too far in the other produces AG. Nevertheless, we suggest two tentative conclusions:

1. The prospects for managing AG with *some* dopaminergic pharmacology, even if it doesn't turn out to be as uniformly specific as a D3 reuptake inhibitor, look increasingly promising.
2. Exploiting the combined power of neurochemical and neuroeconomic insights into AG, some recommended next research steps are obvious. We should be recruiting PG subjects for fMRI studies that look for correlations between the placement of wins in strings of gambles and activation of dopamine neurons in the mesolimbic reward pathway (including PFC). By this method we may hope to both refine our neuroeconomic model of AG, and isolate the brain sites at which neuropharmacological agents should aim.

The review in this section is the basis for our finding it worth hoping that, within a few years, it might be possible for most or all PG to be manageable with a minimally disruptive pill.

³⁰⁴ Baunez, C., Dias, C., Cador, M., & Amalric, M. (2005). The subthalamic nucleus exerts opposite control on cocaine and 'natural' rewards. *Nature Neuroscience* 8: 484-489.

³⁰⁵ Stacchi, F. (2005) Pathological gambling in Parkinson's disease. <http://neurology.thelancet.com> 4: 590-592.

4. Conclusions

EXECUTIVE SUMMARY

There is only one general way to try to make progress with respect to confused and conflicted intuitions on the proper goals and norms of policy around problem gambling. This is to conduct scientific research that attempts to understand what motivates people to gamble, and why some people find it difficult or impossible to gamble within controlled and reasonable limits.

One problem in scientific study of gambling is uncertainty over basic concepts. This uncertainty makes it difficult to coordinate on instruments for recruiting and screening research subjects and interpreting research results. The psychiatric profession has operationalized the concept of 'pathological gambling', but not milder forms of problem gambling that are sometimes thought to be stepping stones to pathological gambling. Another important issue is whether some, many or most pathological gamblers (PGs) are genuinely addicted, in a medical sense, or whether common talk of 'gambling addiction' should be understood as metaphorical.

Based on the BE model of choice over time, we propose an operationalization of disordered gambling and all of its sub-varieties. A 'problem gambler' is anyone who (i) gambles often enough to be consciously aware of trying to develop personal rules to limit their frequency of gambling or their average amount gambled per session or both, but who (ii) regards these rules as inadequately effective because they regret their own gambling behavior on a regular basis, and who (iii) does not meet the psychiatrically accepted (DSM-IV) criteria for pathological gambling (PG). Then a non-addicted disordered gambler (NADG) is any problem gambler or pathological gambler who is not an addictive gambler (AG). A disordered gambler (DG) is anyone who is either a NAGD or an AG.

There is a significant difference between AGs and NADGs that is relevant to treatment strategies and policies. AGs have distinctive personal (neurochemical) properties and are best helped by direct therapy aimed at these properties. NADGs, by contrast, are neuroscientifically normal people who are unusually vulnerable to bundling failures in environments typical of gambling venues. A principal focus of research that aims at reduction of the prevalence of NADG behaviour should be the circumstances under which casino and other standard gambling environments interfere with bundling.

A second class of recommendations from BE models are aimed more directly at NADGs and their significant others, the implementation of which would be a matter for individuals and thus would not entail significant industrial or regulatory reform. The gambling industry may, however, have an interest in participating in and facilitating the operation of an effective counseling system.

The alcohol treatment community and scientific treatment literature are much older and much more extensive than those for NADG. The alcohol treatment system has evolved from a focus on the unitary disease model and very intensive interventions to a focus on a multidimensional biopsychosocial model with a range of interventions that are appropriate for individuals with a range of severity of alcohol problems. Initial indications are that the nascent NADG treatment literature is attending to these developments.

Specific personal-level intervention recommendations based on BE are discussed. First, possible methods to improve NADG's ability to bundle future rewards are recommended.

Second, we recommend several possible ways in which the price of access to gambling opportunities could be increased for NADGs, and several possible ways in which the price of opportunities for social interaction could be decreased for NADGs. Third, we recommend interventions aimed at addressing a particular perceptual distortion among NADGs regarding the value of gambles.

A wide range of behavioural research programmes could contribute significantly to our understanding of disordered gambling, especially among NADGs. Existing work establishes that BE has something to tell us about disordered gambling, but we don't yet know much of what that is. A range of desirable research programmes are briefly described, with remarks on what they could be expected to tell us.

Some, and more probably most or all, PG as operationalized by DSM-IV is in fact addictive gambling (AG). This implies a straightforward policy goal: measures should be taken to try to prevent AGs from accessing commercial or other organized gambling environments. We recommend that responsibility to deny AGs access to gaming establishments should be vested in the operators of the establishments. However, the reasonableness of this recommendation is dependent on resources being provided to make AG identification practical. At present, this condition is not satisfied. The two prospective scientific means of screening for AG, fMRI and brain fluid examinations, are absurdly expensive procedures for application to mass diagnostics outside self-reporting populations.

Two considerations rescue our policy reflections in this area from futility, however. First, evidence indicates that experienced casino staff detect PG customers with high reliability, and that the requisite level of experience is typically acquired quite quickly. Second, if most PGs were not AGs, then the ability of the studies we reviewed to statistically measure AG in populations drawn using screens that over-diagnose PG would be mysterious. Thus there are grounds to suppose that casinos are *already* able to identify and screen out AGs, except perhaps AGs in very early stages of morbidity.

The great hope encouraged by developments in neuroscience of AG is that we will soon be able to exclude AGs from gambling venues in the most benign and welfare-enhancing way of all: by means of a neuropharmacological technology that will reduce or eliminate AGs' interest in gambling in the first place, and their disposition to harm themselves if and when they do find themselves exposed to gambling environments for one reason or another.

However, until the neurochemically specified conditions for AG are determined to general satisfaction in the scientific and clinical communities, and then made the basis for sustained epidemiological research, efforts to estimate the costs to the gaming industry of fully, or nearly fully, managing AG are shots in the dark. Though we are sanguine about neuropharmacological research on AG continuing to be well supported, there is need for *integrated* research that includes BE. The recent breakthroughs in understanding addiction in general and AG in particular rested on the conjunction of basic neurochemistry and neuroeconomics. Neuroeconomics relies heavily on fMRI experiments which in turn require batteries of well-tested behavioral task protocols from BE.

At present there is little capacity for basic fMRI research in South Africa. However, there is capacity for supporting BE work of the kind we have identified as part of the next wave of research that, we hope, will soon turn AG from a serious social problem into a minor nuisance.

4.1. General implications

The most general conclusion of this report is that disordered gambling (DG) divides, for scientific and treatment / intervention purposes, into two fundamental kinds. First there is *non-addictive disordered gambling* (NADG). We propose that NADGs be identified by criteria motivated by Behavioral Economics (BE). In particular, a NADG is someone who is not an addictive gambler (AG) and who (i) gambles often enough to be consciously aware of trying to develop personal rules to limit their frequency of gambling or their average amount gambled per session or both, but who (ii) regards these rules as inadequately effective because they regret their own gambling behavior on a regular basis. NADGs are in turn divided into two groups: (a) those who are diagnosed as 'pathological gamblers' (PGs) on the basis of the current or future updated DSM criteria and (b) 'problem gamblers', who meet criteria (i) and (ii) but not (a).

Leading our recommendations with this classification issue may appear fussy and academic (in a pejorative sense) to some. We emphasize it because we think it unlikely that efficient progress on improved intervention against DG will happen unless scientific classification and clinical classification are brought into line with one another. This is for two reasons:

- (1) Subjects for scientific research must often be recruited from clinical populations, using screens developed for clinical purposes.
- (2) Research on DG is, to a large extent, applied research with intended practical outcomes. It is difficult to communicate its relevance to the clinical and policy communities in the absence of shared nomenclature.

Note that our proposed classification is relatively conservative. It leaves undisturbed the current operationalization of PG as established by DSM-IV and reflected in most existing PG screens. The novelties in our proposal are two:

- (I) We use criteria derived from BE to distinguish DGs from people who like to gamble and do not have recurrent problems. No scientifically based criteria have previously been proposed for doing this. Insofar as DG is a motivational disorder, as we have argued that it is, then it is appropriate that the science that rigourously classifies it should be the science of behavioural motivation, choice and reward – BE.
- (II) We emphasize the distinction between AG and NADG. This reflects our belief, justified and explained in Chapter 3, that there is a genuine phenomenon of gambling addiction, in the sense that AGs have distinctive, neurochemically and neurodynamically specifiable and diagnosable properties.

(II) can be cited to illustrate our point about the importance of bringing scientific and clinical nomenclature into agreement. It is doubtful that the kinds of behavioral interventions we recommend as responses to NADG will be optimal for many AGs. Note our use of 'optimal' rather than 'effective' here. Most addicts, including AGs, ultimately recover without resort to neuropharmacological therapy. Such people hold relapse at bay by successful development and maintenance of personal rules (see below). However, most also go through several episodes of relapse on their way to recovery, and these can be catastrophically costly or even fatal (with suicide being the main proximate cause of death in fatal AG). As we will discuss in section 4.3 below, we expect that there will shortly

be effective neuropharmacological interventions for AG. *If* their side-effects are minimal and/or readily manageable, then these should be the standard first response to AG. In our view, the state and the commercial gaming industry should cooperate to attempt to minimize access to commercial gambling by untreated AGs. We will return in section 4.3 below to questions about practical mechanisms for seeking to do this. We note that it is already the practice of most legal casinos, including those in South Africa, to deny service to PGs. Since we believe that most PGs are in fact AGs, our proposal simply adds the promise of additional and more exact diagnostic methods for implementing what is already a prevailing norm with existing policy justified by it.

Policy issues around reduction of NADG behaviour are more complicated. Evidence and theoretical considerations reviewed in this report suggest that NADGs will not share stereotypical distinctive personal properties. Many people might be vulnerable to becoming NADGs in some kinds of environments but would not be at significant risk in others. Furthermore, by our proposed operationalization above, being a NADG (or not) is partly a function of subjective preference. If a person wishes to spend what others regard as a disproportionate share of their income gambling, if they make no efforts to erect personal rules to protect longer-term interests from costs of gambling, and if they are not diagnosable as AGs, then – putting aside issues that may arise if they have children in whose welfare there is a legitimate and accepted public interest – it seems there is no basis for undertaking intervention to modify their behaviour except paternalism. Such paternalism would appear to be inconsistent with the general social and legal norms of the South African democratic order established since 1994.

Note that what we just said implies no opinion, one way or the other, on the issues raised by Howard Rachlin's contention (see Chapter 1) that the state should not derive significant revenue from commercial gambling. That is ultimately a question of normative political philosophy that is beyond the brief of this report.

Closely related to this question, but not quite so obviously beyond the remit of our attention as scientists, is the issue of whether the public benefits of legal but regulated gambling do or do not exceed the public costs. A recent comprehensive study by economist Earl Grinols³⁰⁶ argues that, in the case of the United States, public costs exceed public benefits. We will offer no opinion on this conclusion, or seek to consider the probability that it extends to South Africa. However, we will relate the estimations of Professor Grinols's model to our recommendations on management of DG to possible costs of these measures in terms of industry and state revenues.

If NADGs cannot be distinguished from other gambling services customers except retrospectively (i.e., by reference to whether their preferences reversed twice according to the pattern for bundling attempt and failure identified in Chapter 1), then policy response to NADG cannot consist in trying to restrict their access to commercial gaming establishments. Instead, policy must have as its aim control of the extent to which gamblers' abilities to successfully bundle in accordance with their personal rules are deliberately subverted.

This is a quite subtle issue. Casinos and other commercial gaming establishments are essentially environments designed to encourage gambling. Were they required to design their environments so as to make gambling appear to be positively unpleasant – as some liquor stores in (e.g.) Scandinavian countries are required to do with respect to drinking –

³⁰⁶ Grinols, E. (2004). *Gambling in America: Costs and Benefits*. Cambridge: Cambridge University Press.

this would impose significant costs *both* on the gaming industry and on non-disordered gambling customers.

The only justification we can imagine for such a policy would be if it were a second-best option chosen in the wake of a decision that it would be best to have no legal gambling at all, but for the opportunities outright bans present to organized crime. This is clearly not the prevailing policy attitude in South Africa.

Therefore, we assume that casinos should legitimately aim to make their environments comfortable, and indeed lively, cheerful and fun. But BE evidence clearly indicates that ‘lively’ and ‘cheerful’ denote bundles of environmental variables that make it harder for people to bundle. This leads us to a very general point. The moment a society decides to have legal, regulated gambling, it thereby decides to tolerate a level of NADG greater than zero. Any casino environment not designed to be grim and aversive will interfere with *some people’s* capacities to bundle. Deciding how much NADG behavior is a cost worth bearing in exchange for the benefits of legal access to gambling in pleasant surroundings is a normative social choice on which different societies will have different cultural preferences.

We note that Professor Grinols’s model estimates that in America what he calls ‘problem gambling’, meaning DG that does not meet the criteria for PG, accounts for approximately 30% of casino revenues. If this estimate is in error, it probably errs on the low side. For reasons discussed in Chapter 1, screens based on DSM criteria tend to overclassify problem gamblers as PGs. Thus some gamblers identified in studies that contribute to Professor Grinols’s data as PGs likely should have been counted in the data on problem gamblers. Notice that this is *good news* for the gaming industry insofar as there is agreement with our suggestions that most PGs are AGs and that AGs should be excluded as far as is possible from commercial gambling opportunities. The higher the ratio of problem to pathological gamblers in the population of DGs, the better this is for industry revenues in a context where serious efforts are undertaken to reduce DG prevalence. After all, sound policy should aim to deprive the industry of 100% of the revenues it collects from AGs.

Let us then take Professor Grinols’s estimate of 30% as an *upper bound* on the casino revenue reduction that would result from *complete* eradication of NADG. We have argued that this is not a reasonable goal in present-day South Africa. So how far below the 30% upper bound would costs to industry be of modifications to their environments that facilitated better bundling by patrons? This question admits of no answer – not even a principled *estimate* – in the absence of the sort of BE scientific research we will discuss in section 4.3 below. We do not know to what quantitative extent drinking alcohol, or having the body’s natural clock disrupted by manipulated ambient light, or being smiled at by attractive young casino staff, or hearing coins jingling in trays, or being surrounded by flashing lights, or etc., impair bundling capacities in different groups of people. To what extent are differences among people in these respects functions of age, gender, wealth, level of education or the presence of family and friends? *These are all things we will need to try to find out if industry and government wish to be able to project the impact of possible casino environment regulations on revenues.* They are all things BE researchers know how to find out given resources for research.

We thus recommend that responsible authorities in government, industry and the clinical community organize funding of BE research into the effects of casino environment variables on capacities to bundle choices. We will say more about this in section 4.3 below.

First, however, let us say a bit about what we *do* know already that is relevant to possible institutional-level policy (as opposed to individual-level treatment interventions) in connection with NADGs.

We will first consider suggestions raised by Ainslie's model of bundling, and then focus on the specific advice that would be added by Rachlin. Before doing either, however, we should reiterate that all that is being attempted at this point is logical extrapolation from theoretical perspectives.

There are two general classes of interventions to facilitate bundling that, in the case of casino gambling, imply changes to the design of betting rules, and a third that implies modifying the design of casinos. First, Ainslie's model suggests that DGs should be helped to find ways of committing to whole sequences of gambles at once, instead of anteing many small, separate stakes. This can help the heavily discounted utility of later gambles stack up against the utility of looming gambles in influencing the person's current decision.³⁰⁷ Second, DGs should be helped to find ways of imposing delays between commitments of stakes and receipt of outcomes. This tactic would increase the probability of a DG's making a decision at that point of the reward value curve (see Chapter 1, Figure 2) when the larger later reward (LLR) is preferred over the sooner smaller reward (SSR).³⁰⁸ (In light of this point, we can note that on Ainslie's model lotteries will be the least dangerous form of gambling, an idea supported by empirical research in most jurisdictions – with South Africa being the interesting exception.) Third, the LLR value will typically be boosted to the extent that it is actively represented in the present environment. The extent of DG could be reduced by reminders in casinos regarding what is potentially being given up by the gambler. For example, casinos could be required to post signs that asked, "What would you do next week with the money you'll lose today?" or, "Could the money you'll lose today be used to buy clothes for your kids next week?"³⁰⁹ Such reminders would emphasize that gambling represents a choice and that there are opportunity costs of gambling. Of course, this idea begins to take us in the direction of making casino environments less pleasant for non-disordered gaming service customers. As we have emphasized, this is a trade-off issue that calls for normative social choice. It also raises the cost-benefit questions that BE could answer in principle but on which the necessary research remains to be done.

³⁰⁷ Although not undertaken from the BE perspective, some research has shown that betting machines modified to accept reduced maximum bet amounts were played for shorter durations, and were associated with smaller losses, fewer individual bets, and lower levels of alcohol consumption and smoking. (Sharp, L., Walker, M., Coughlan, M., Enerson, K., and Blaszczynski, A. (2005). Structural changes to electronic gaming machines as effective harm minimization strategies for non-problem and problem gamblers. *Journal of Studies on Gambling* 21: 503-520.) This likely represents effective direct manipulation of the mesolimbic reward system as discussed in Chapter 3, rather than an intervention to promote bundling or other improved bookkeeping.

³⁰⁸ Note, however, that the neuroeconomic and neurochemical models of the reward system reviewed in Chapter 3 tell us that the main reward for the gambler is not the payout but the initiation of the assurance of a surprise. Nevertheless, the longer into the future the resolution of surprise is pushed, the weaker the cue for reward constituted by initiating a bet. So the general point made in the text holds, albeit with an important qualification.

³⁰⁹ Even if some or many casinos would post such signs voluntarily, regulation of such measures, if they are deemed normatively appropriate in the first place, is preferable so that casino operators can be confident that their competitors will not try to free ride on the policy.

4.2. Personal- and community-level interventions recommended by BE

General treatment issues

A second class of recommendations from BE models are aimed more directly at NADGs and their significant others. The implementation of these would be a matter for individuals and thus would not entail significant industrial or regulatory reform.

The gambling industry may, however, have an interest in participating in and facilitating the operation of an effective counseling system, for which the measures discussed in this section are directly relevant. Regardless of whether a response to DG stems from application of BE, it will be undertaken as part of a larger intervention system. So, before discussing the specific BE recommendations we will first comment generally on the current state of knowledge regarding treatments for NADG.

A DG psychological treatment literature based on scientific evidence is only in the nascent stage of development, although it is growing rapidly. A review of DG treatment studies published in 2003³¹⁰ included 11 studies, while a similar review published in 2005³¹¹ included 22 studies. The 2003 review concluded that cognitive-behavioral (CB) interventions seemed to have the most promise, perhaps because CB is the dominant paradigm in clinical psychology, while the 2005 review concluded only that “psychological treatments” were better than no treatment. Both of these reviews excluded a large number of other studies because they contained critical methodological flaws. Working out the numerous methodological kinks is quite common in the incipient states of any scientific literature, including treatment for DG³¹². Among the most common methodological problems were lack of random assignment to treatment and control conditions, inadequate or inadequately specified assessment procedures, inadequate or inadequately specified treatment protocols, ill defined and/or poorly measured treatment outcomes, and insufficiently lengthy follow-up periods. Clearly, strong recommendations for specific NADG treatments would be premature at this time, given the current state of the scientific literature.

Given that an active alcohol treatment community has existed since the 19th century, at least in the US³¹³, it will be instructive to summarize some developments in that community with an eye towards their possible relevance to NADG. Up through the late 20th century, alcohol treatment was dominated by the disease model, which maintained that alcoholism was a unitary disorder, that an individual with alcohol problems less severe than alcoholism was merely on his way to becoming an alcoholic, and that recovery required intensive (e.g., inpatient hospitalization) and extensive (e.g., long duration) interventions. However, these beliefs among treatment providers were developed essentially without the

³¹⁰ Toneatto, T., and Ladoceur, R. (2003). Treatment of pathological gambling: a critical review of the literature. *Psychology of Addictive Behaviors* 17: 284-292.

³¹¹ Pallesen, S., Mitsem, M., Kvale, G., Johnsen, B.H., and Molde, H. (2005). Outcome of psychological treatments of pathological gambling: a review and meta-analysis. *Addiction* 100: 1412-1422.

³¹² Blaszczynski, A. (2005). Conceptual and methodological issues in treatment outcome research. *Journal of Gambling Studies* 21: 5-11; Nathan, P.E. (2005). Methodological problems in research on treatments for pathological gambling. *Journal of Gambling Studies* 21: 109-114.

³¹³ Tucker, J.A. (1999). Changing addictive behavior: Historical and contemporary perspectives (pp. 3-44). In Tucker, J., Donovan, D., and Marlatt, G. (Eds.), *Changing addictive behavior: Bridging clinical and public health strategies*. New York: Guilford.

benefit of scientific evidence. It is not surprising, then, that when alcohol treatment research blossomed during the last 25 years of the 20th century, many of the beliefs dictated by the disease model turned out not to be the most effective means to deal with alcohol problems.

The contemporary perspective in the treatment community³¹⁴ is that alcohol problems are multidimensional rather than unitary, exist on a continuum of severity (and less severe cases do not inevitably become more severe cases), and that a truly effective intervention system will include a range of prevention³¹⁵ and treatment interventions that are appropriate for the range of severity levels³¹⁶. The current emphasis is on the integration of both clinical and public health approaches to alcohol problems³¹⁷, which is similar to the development of programs to deal with tobacco use³¹⁸.

Interventions lie on a continuum of intensity. At the more intense end of the continuum of alcohol treatments would be a several-week inpatient hospitalization or a several-month residential program. Such programs are very expensive, affect a relatively small number of people, and should be reserved for the most serious and recalcitrant cases. At the less intense end of the continuum would be a universal or selective prevention program that reached the entire or selected segment of the population. Such programs are much less expensive (per person affected by the program) and reach many more people than more intense interventions. The more and less intense parts of the continuum correspond to clinical and public health interventions, respectively. Clinical interventions are labour intensive (e.g., one counselor treats one client over time), expensive, and must be highly selective regarding whom they treat. Public health interventions are less labour intensive (e.g., one public health professional could affect tens or hundreds of persons), cheaper, and can be less selective regarding whom they treat. An integrated clinical and public health intervention system would allocate its resources to interventions along the continuum according to the distribution of types of cases in the population, the extent of

³¹⁴ See, for example, Hester, R.K., and Miller, W.R (2002). *Handbook of Alcoholism Treatment Approaches* (3rd Edition). New York: Allyn & Bacon.

³¹⁵ Prevention programs are conceptualized according to the segment of the population they target: *Universal* programs are directed at the entire population. For example, a media campaign, which would contact most of the population, that emphasized the risks associated with excessive gambling would be a universal prevention effort. *Selective* programs are directed at individuals believed to be at risk for developing the disorder, but before they show any symptoms. In Chapter 2 we noted that preliminary evidence indicates that early adolescents high in impulsiveness are at increased risk to become DGs later. If this turns out to be a reliable finding, then high impulsiveness adolescents could be identified and targeted with a DG prevention program. *Indicated* programs are directed at individuals showing early signs of the disorder, but before they are severe enough to be actually diagnosable. Institute of Medicine (Mrazek, P.J., and Haggerty, R.J. [Eds.]). (1994). *Reducing risks for mental disorders: Frontiers for preventative intervention research*. Washington, DC: National Academy Press.

³¹⁶ Institute of Medicine (1990). *Broadening the base of treatment for alcohol problems*. Washington, DC: National Academy Press.

³¹⁷ In Tucker, J.A., Donovan, D., and Marlatt, G.A. (Eds.), *Changing addictive behavior: Bridging clinical and public health strategies*. New York: Guilford.

³¹⁸ Abrams, D.B, Emmons, K.M, Niaura, R., Goldstein, M.G., and Sherman, C.B. (1991). Tobacco dependence: An integration of individual and public health perspectives (pp. 391-436). In Nathan, P., Langenbucher, J., McCrady, B., and Frankenstein, W. (Eds.), *Annual review of addictions research and treatment*. New York: Pergamon.

harm caused by the disorder in the population, and the outcomes the system hoped to achieve. Such integration is much more efficient and effective than if the various types of interventions are managed independently of one another.

Three recent developments in alcohol treatment may be particularly relevant to NADG. First, most people who recover from alcohol problems do so without the aid of any formal intervention³¹⁹, so a current priority for research is to better understand the *natural recovery* process so that that information can be incorporated into treatment interventions³²⁰. Natural recovery appears to be quite common among NADGs as well³²¹. Second, for many people with alcohol problems, the particular type of treatment intervention they receive is much less important than whether or not they are properly motivated to change their drinking behavior³²². As contended in section 4.1 above, since BE basically is the scientific study of motivation, BE should rise to a more prominent position in treatment research. Third, many people with alcohol problems can undergo significant change as a result of *brief interventions*, which consist of only one or two sessions with a counselor³²³. Widespread implementation of such brief interventions will reach a much larger segment of the population of individuals with alcohol problems.

The importance of those hard-won lessons in alcohol treatment have not been ignored by some in the DG treatment community. For example, in addition to the typical, relatively intense DG treatments³²⁴, DG intervention researchers have begun to develop universal prevention programs³²⁵, indicated prevention programs³²⁶, and brief interventions³²⁷.

-
- ³¹⁹ Sobell, L.C., Cunningham, J.A., and Sobell, M.B. (1996). Recovery from alcohol problems with and without treatment: Prevalence in two population studies. *American Journal of Public Health* 86: 966-972.
- ³²⁰ Tucker, J.A., Vuchinich, R.E., and Rippens, P.D. (2002). Predicting natural resolution of alcohol-related problems: A prospective behavioral economic analysis. *Experimental and Clinical Psychopharmacology*, 10: 248-257.
- ³²¹ Abbott, M.W., Williams, M.M., and Volberg, R.A. (2004). A prospective study of problem and regular nonproblem gamblers living in the community. *Substance Use and Misuse* 39: 855-884.
- ³²² Miller, W.R., Rollnick, S., and Conforti, K. (2002). *Motivational interviewing: Preparing people for change* (2nd Edition). New York: Guilford.
- ³²³ Miller and colleagues, in the source cited by the previous note.
- ³²⁴ Toneatto, T., and Ladoceur, R. (2003). Treatment of pathological gambling: a critical review of the literature. *Psychology of Addictive Behaviors* 17: 284-292; Pallesen, S., Mitsem, M., Kvale, G., Johnsen, B.H., and Molde, H. (2005). Outcome of psychological treatments of pathological gambling: a review and meta-analysis. *Addiction* 100: 1412-1422.
- ³²⁵ Bryan, A.M., Dickson, L., Derevensky, J.L., Gupta, R., and Lussier, L. (2005). The application of youth substance use media campaigns to gambling: A critical examination. *Journal of Health Communications* 10: 681-700; Messerlian, C., Derevensky, J., and Gupta, R. (2005). Youth gambling problems: a public health perspective. *Health Promotion Initiatives* 20: 69-79; Ladouceur, R., Vezina, L., Jacques, C., and Ferland, F. (2000). Does a brochure about pathological gambling provide new information? *Journal of Gambling Studies* 16: 103-107.
- ³²⁶ Takushi, R.Y., Neighbors, C., Larimer, M.E., Lostutter, T.W., Cronce, J.M., and Marlatt, G.A. (2004). Indicated prevention of problem gambling among college students. *Journal of Gambling Studies* 20: 83-93.

In one universal program, for example, Bryan and colleagues distributed brochures containing information on DG in malls and other commercial establishments, and showed that knowledge about DG increased among persons in those areas where the brochures were distributed. In an indicated prevention program, Takushi and colleagues identified college students who were engaging in high-risk gambling, and provided them with cognitive-behavioral skills-training, motivational interviewing, cognitive correction, discussion of gambling consequences, and relapse prevention techniques. In Hodgins's and colleagues' brief intervention, identified DGs were given a self-help DG workbook and a motivational interviewing telephone call, and the treated DGs were significantly improved at a 24-month follow-up relative to the control group.

The efficacy and effectiveness³²⁸ of these various DG interventions will be evaluated in future research. But the relevant point for present purposes is that the DG community, at least in some jurisdictions, apparently has learned that a truly effective intervention system will provide a diverse set of responses that are appropriate for the range of manifestations of DG, and that the intervention system should be arranged in a stepped care fashion³²⁹. Stepped care refers to a treatment system in which individuals receive interventions at an intensity level appropriate for the severity of their symptoms, and only move on to more intensive interventions if they don't improve. An idealized stepped care intervention system is illustrated in Figure 1. This hypothetical system contains four interventions, which increase in intensity from left to right. Persons entering the system are assessed with evidence-based procedures. Based on the results of the assessment and clinical judgment, cases are referred to the treatment with an intensity-level appropriate for their situation. Initial referral decisions are based on, for example, severity of the symptoms (with more severe cases going to more intense treatments) and social and economic circumstances (e.g., employment status, social support network, ability to pay out-of-pocket or with insurance). Regardless of the initial referral decision, each case will have a positive or negative outcome from the initial treatment. Cases with initial negative outcomes would then be referred to a more intense treatment, where they would again have either a positive or negative outcome. Cases with initial positive outcomes are discharged from treatment and are monitored occasionally to assess their ongoing status. If an initially successfully treated case relapses, they would reenter the system and again be assessed and referred to treatment. This hypothetical system is idealized, to be sure, but it illustrates how an intervention system can efficiently allocate resources and produce the most positive outcomes possible relative to limited resources.

³²⁷ Hodgins, D.C. (2005). Implications of a brief intervention trial for problem gambling for future outcome research. *Journal of Gambling Studies* 21: 13-19; Hodgins, D.C., Curry, S., el-Guebaly, N., and Peden, N. (2004). Brief motivational treatment for problem gambling: a 24-month follow-up. *Psychology of Addictive Behaviors* 18: 293-296.

³²⁸ *Efficacy* refers to whether or not a treatment has been shown to produce better clinical outcomes when compared to a control group in a methodologically sound study. *Effectiveness* refers to whether or not a treatment will maintain its efficacy when implemented under normal clinical conditions.

³²⁹ See, for example, Sobell, M.B., and Sobell, L.C. (1999). Stepped care for alcohol problems: An efficient method for planning and delivering clinical services (pp. 331-343). In Tucker, J., Donovan, D., and Marlatt, G. (Eds.), *Changing addictive behavior: Bridging clinical and public health strategies*. New York: Guilford.

4.3. Future behavioral economic research and NADGs

While the preceding chapters will have shown that behavioral economic approaches contribute significantly to our understanding of substance abuse and disordered gambling, there is still much more to be found out. Here we briefly indicate some of the lines of research that could most be expected to contribute to improved understanding and management of NADGs.

Although the preceding chapter may have made it seem as though the only sort of research moving forward in this area is neuroeconomics, or at least some combination of branches of neuroscience, this was not our intention. Neuroscience experiments shed light on questions about how decisions are made when subjects perform well-designed tasks, and the project of determining what sorts of tasks are appropriately discriminating is largely behavioral. Thus, over and above the reasons indicated in the preceding section, there is reason to invest in further specifically behavioral study of disordered gambling. If our principal claim regarding the two main types of disordered gambling – addicted gambling (AG), and non-addicted disordered gambling (NADG) – is correct, then this is especially so in the case of NADGs. These people won't as a group share a distinctive brain property like addicted gamblers, and their vulnerability to bouts of destructive gambling may turn out to have more to do with environmental factors than anything else. Behavioral research can do a great deal to determine the extent to which this is so, and thereby direct the attention of policy discussion to the issues that really do bear in prevalence, vulnerability and effective treatment.

First, though, before considering the relationship between vulnerability to NADG and environmental factors, comes some more basic research. As Chapter 2 made clear, BE approaches have some success in separating disordered gamblers from non-disordered ones. The research conducted to date on disordered gamblers is, though, far less extensive than that on abusers of and addicts to various substances. Also, most of the research focuses on impulsivity in the specific sense of intolerance for delays, while relatively little research to date has concerned itself with tolerance to variations in the probability of reward.

The sort of research suggested is, fairly obviously, empirical study involving populations of individuals with varying gambling severity ranging from recreational non-disordered gamblers, through gamblers with moderate and severe disordered gambling (as measured by some suitable screen).³³⁰ Such a study should also include comparison subjects with a consumption disorder other than gambling.³³¹ The extent to which disordered gamblers of varying severity differ in respect of their discounting for delays and probabilities compared to normal controls and substance users will thereby be made known.

The resulting knowledge of the extent to which NADGs of varying gambling severity were more impulsive in the case of delays or reduced probabilities would be a significant step in the direction of a more precise behavioral characterising of the phenomenon of disordered gambling than we currently have. We'd know better than we currently do in what specific ways NADGs differ from those who gamble recreationally and non-destructively.

³³⁰ Too little research on disordered gambling includes samples of subjects with enough variation in gambling severity. Some difficulties resulting from this, in the interpretation of the possibly conflicting results of studies each with subject pools of quite different gambling severity, were discussed in Chapter 2.

³³¹ We could specify various exotic technical details of possible experimental designs here, but doing so would mean departing from the proper business of this survey.

Whatever is discovered will be relevant to the identification of disordered gamblers, and to the management and treatment of NADGs.

A complementary form of research would be *longitudinal* studies in which populations of subjects were assessed behaviorally at a variety of ages, in order to establish the extent to which impulsiveness at one age is a predictor of disordered gambling at others. As was mentioned earlier, impulsiveness at age 12-14 by the criteria of a psychiatric instrument has been found to be a predictor of disordered gambling at ages 15-17.³³² We know that individual discount rates vary widely, but we do not yet know enough about the extent to which they are stable over extended periods in the life of developing people, and hence the extent to which early assessment using discounting tasks could be a useful way of identifying those vulnerable to a range of consumption problems including disordered gambling. This sort of research could begin to deliver results within as little as two years, but as long as the initial cohort of subjects was large enough, and carefully enough tracked over time,³³³ it could continue to lead to further discoveries for an extended period.

The kind of work suggested above, involving measurement of discounting properties as relatively 'static' features of individual subjects, though, would be insufficient by itself. As was also explained in Chapter 2, we know of a variety of ways in which cues in the environment, or other consumption activities such as consuming alcohol, can lead to transient changes in discounting, or in performance at other decision tasks. For NADG subjects, it is probable that these sorts of influences are a significant contributor to disordered gambling. Again, behavioural research should be able to tell us a great deal more about these influences and changes, by constructing experimental settings in which they can be observed, and finding ways of studying them in natural settings. We emphasise that we are here discussing *behavioural* research, which is to say that the point of the experiments would be to track the economic behaviour (that is responses to changes of price, reward magnitudes, etc.) of subjects under specific conditions.

We can distinguish three broad types of behavioural research into changes in discounting:

First is longitudinal work with considerably shorter delays between successive assessments of the subjects. This sort of work should also determine what significant life events of various sorts had taken place over the course of the study. We know anecdotally, from the reports of clinicians, and from some empirical research, that relapses by addicts are more likely following events such as separations, bereavement, and retrenchment.³³⁴ We could, though, know much more about the ways in which life events are associated with variations in impulsivity. Since people with heightened impulsivity do damaging things besides excessive gambling, including driving dangerously and engaging in risky sexual behaviours, such research would be of wide-ranging significance.

Second is experimental work oriented to determining what sorts of environmental cues (either by their presence or their absence) other consumption (including of alcohol) impact on discounting with a view to better understanding determinants of vulnerability.

³³² Vitaro, F., Arseneault, L. and Richard E. Tremblay, R. E. (1999). Impulsivity predicts problem gambling in low SES adolescent males. *Addiction*, 94(4), 565-575.

³³³ This is, large enough that, even with inevitable attrition from subjects migrating or otherwise becoming impossible to contact or unavailable for follow-up work, the cohort would still permit statistically significant results to be reported in later years.

³³⁴ Vuchinich, R.E., and Tucker, J.A. (1996). Alcoholic relapse, life events, and behavioral theories of choice: A prospective analysis. *Experimental and Clinical Psychopharmacology* 4: 19-28.

In particular this work should focus on environmental cues associated with the lives of populations known to be vulnerable to disordered gambling, and ones associated with gambling venues. The latter need not be cues uniquely associated with gambling venues: it is likely that other controlled environments such as bars and shopping centres with an interest in leading people to spend more share some of the same key features, such as:

- obscuring the outside world so as to disrupt accurate perception by customers of how long they have been gambling, drinking or shopping,
- prominently displaying desirable luxury objects clearly identified as in some sense accessible to passers-by (for example by being the prizes in currently running contests),
- organising for the presence of attractive people, or at least images of them,
- orchestrating lively events drawing and entertaining crowds of people to create and maintain a sense of activity and excitement,
- looking like expensive places.

Third, and to a significant extent overlapping with the previous area, is experimental work oriented to determining what sorts of environmental cues (either by their presence or their absence) impact on discounting with a view to better therapeutic intervention. This includes the sort of work noted as desirable in the previous section, which is work on how better to promote ‘bundling’ of choices in subjects with difficulties in doing so. Again, in a sense we already know that some of these things can make a difference, for example that ‘thinking about your children’ can help someone view the ‘one’ offered cigarette they are thinking of accepting as representing the risk of a long period of cumulatively expensive consumption, and thereby refuse it. Similarly the practice of organising the presence of attractive women in so many attempts to promote motor cars is not *mysterious* to us, any more than that people make worse decisions when they have been drinking. What we lack, though, is a more clearly integrated theory and body of empirical work on the properties and magnitudes of the changes brought about by various cues and triggers. So, to refer again to a study discussed in Chapter 2, we now have behavioural research showing that contemplating attractive women leads to temporarily increased sensitivity to delay in male subjects.³³⁵ This is a more specific and quantitatively precise description of what we (or advertisers) knew already, so that along with other work of the same sort, we can work out where to focus our attention when trying to direct policy and therapy.

Next, although the preceding paragraphs have focussed to a large degree on discounting, it is worth remembering that behavioral economics, as explained in chapter 1, has much more to it than study of discounting. Consequently the tools of BE may be able to do more still to help us understand the differences, from the perspective of concern with disordered gambling, between different types of gambling, and relations between gambling and other activities. So, to recap two points made in chapter 1:

- Given that different games of chance vary in their potential to generate disordered gambling, demand curve studies that compared demand for gambling opportunities across games could yield valuable information regarding the games’ individual “abuse liability.”

³³⁵ Wilson, M. and Daly, M. (2004). Do pretty women inspire men to discount the future? *Proceedings of the Royal Society of London B (Suppl.)*, 271, S177–S179.

- Since we know that the variety and availability of valuable activities other than substance use can have a critical influence on substance consumption, we should attempt to establish whether similar relations exist between changes in disordered gambling and alternative activities. In fact, there is not really any doubt about the 'whether'; the question is *in what sorts of ways* the relations hold in the case of gambling. (It would be interesting to have, for example, household surveys and subsequent microeconomic analyses that tested the effect of new sports facilities or entertainment complexes in communities where there were already organized gambling venues.)

Finally, to return to a discussion in Chapter 2, we need to know more than we do about how the main research instruments of behavioural economics and cognitive neuroscience relate to one another. In particular, a study in which subjects of varying gambling severity completed both delay and probability discounting procedures and the Iowa Gambling Task (IGT) or one of its variants would go a long way to working out to what extent the findings of the two approaches are different ways of finding the same thing (i.e. both approaches agree on who the disordered gamblers are in the sense that behavioural impulsivity and poor IGT performance both effectively separate disordered gamblers into one pile, and others into a second pile) or are instead complementary in tracking importantly different aspects of disordered gambling.

4.4. Interventions for AG

In Chapter 3 we presented the impressive evidence for the claim that some, and more probably most or all, PG as operationalized by DSM-IV is in fact addictive gambling (AG). In our view this implies a straightforward policy goal: measures should be taken to try to prevent AGs from accessing commercial or other organized gambling environments. AGs suffer from a diagnosable and distinctive neurochemical disorder that can be life-destroying. The disorder directly threatens the welfare of their families, friends and business colleagues. It is more or less inconceivable that the social benefits of allowing AGs to access dedicated gambling environments could exceed the social costs. Furthermore, the state's interposing itself between AGs and their target of addiction is consistent with existing policies in analogous domains.

We should stress that we do not favour punishing identified AGs who attempt to access gaming establishments with criminal or civil sanctions. It is perhaps not our station as scientists to point out that this would be inconsistent with the strong rights-based constitutional culture of democratic South Africa. However, it clearly *is* within our brief to point out that trying to control AG behaviour by threat of punishment would almost certainly be inefficacious. As we noted in Chapter 3, the reward system, the biological mechanism that addiction comes to control, does not strongly respond to punishment. In general, neuroscience knows very little about how punishment responses are mediated in the brain. We know *just* enough to know that we should be very cautious in relying on it for any important social purposes except, in some cases (but not the one at hand) to deter third parties from selfish actions.

We therefore recommend that responsibility to deny AGs access to gaming establishments should be vested in the operators of the establishments. We are not legal experts and will not comment on details of possible sanctions for non-compliance with this responsibility. We will merely note, as economists, that sanctions must clearly be costly enough to exceed revenue benefits that casinos might anticipate from serving AGs.

The reasonableness of the above recommendation is absolutely dependent on resources being provided to make AG identification practical. At present, this condition is not satisfied. Examination of a gambler's brain under fMRI is not sufficient, at this stage of knowledge, to reliably identify AGs. A more immediately servicable method is testing cerebrospinal fluid (CSF) dopamine levels – though it must be added that this method alone would produce many false positives (probably more false positives than accurate diagnoses), and so would have to be relied on only in the case of an independent diagnosis of PG based on a stringent DSM-derived screen (e.g., the Canadian Gambling Index). The same qualification certainly applies to fMRI screening at this point, though we expect that this might change as fMRI technology, clinical experience with fMRI and scientific knowledge based on neuroimaging all improve. But to some extent all of these considerations are academic, because both fMRI and CSF examinations are absurdly expensive procedures for application to mass diagnostics outside self-reporting populations.

Two considerations rescue our policy reflections in this area from futility, however. First, evidence indicates that experienced casino staff detect PG customers with high reliability, and that the requisite level of experience is typically acquired quite quickly. Second, we argued in Chapter 3 that if most PGs were not AGs, then the ability of the studies we reviewed to statistically measure AG in populations drawn using screens like SOGS that over-diagnose PG would be mysterious. Thus there are grounds to suppose that casinos are *already* able to identify and screen out AGs, except perhaps AGs in very early stages of morbidity. The existence of diagnostic tools that are impractical *en masse* should encourage casinos – or encourage the state to encourage casinos – to adopt more rather than less stringent exclusionary practices. After all, if a person believes they have been misidentified as a AG by a casino, it is possible (and should be possible as a standardly insured medical procedure) for them to obtain an expert diagnosis which, if they are right, would gain them re-admission.

Of course, the great hope encouraged by the developments reported in Chapter 3, section 6, is that we will soon be able to exclude AGs from gambling venues in the most benign and welfare-enhancing way of all: by means of a neuropharmacological technology that will reduce or eliminate AGs' interest in gambling in the first place, and their disposition to harm themselves if and when they do find themselves exposed to gambling environments for one reason or another.

The only significant cost to attempting to prevent AGs from gambling in commercial venues, aside from enforcement costs that are bound to be low on the margin in any already heavily regulated and inspected industry, are borne by gaming enterprises. These costs may be far from trivial.³³⁶ From Professor Grinols's model, discussed in section 4.1, we have derived the estimate that if 80% of PGs diagnosed in the United States are AGs, and all PGs who presently gamble in legal US casinos were successfully excluded, this would reduce average casino revenues by 25%. Now, the figure of 80% is highly arbitrary. Though we have argued that most PGs are probably AGs, we also pointed out in section 4.1 that the surveys on which prevalence models are forced to rely systematically overestimate the proportion of PGs in populations of DGs.

³³⁶ It should be pointed out that in trying to calculate costs of AG exclusion to the gaming industry, one should not neglect to consider the fact that successful such exclusion, especially if achieved by neuropharmacological management of AG, might significantly reduce popular suspicion of and, in some quarters, outright hostility to the industry. This must be entered into the ledger on the opposite side to the projected costs. We believe that industry managers do not neglect this point.

In general, until the neurochemically specified conditions for AG are determined to general satisfaction in the scientific and clinical communities, and then made the basis for sustained epidemiological research, efforts to estimate the costs of fully, or nearly fully, managing AG are shots in the dark.

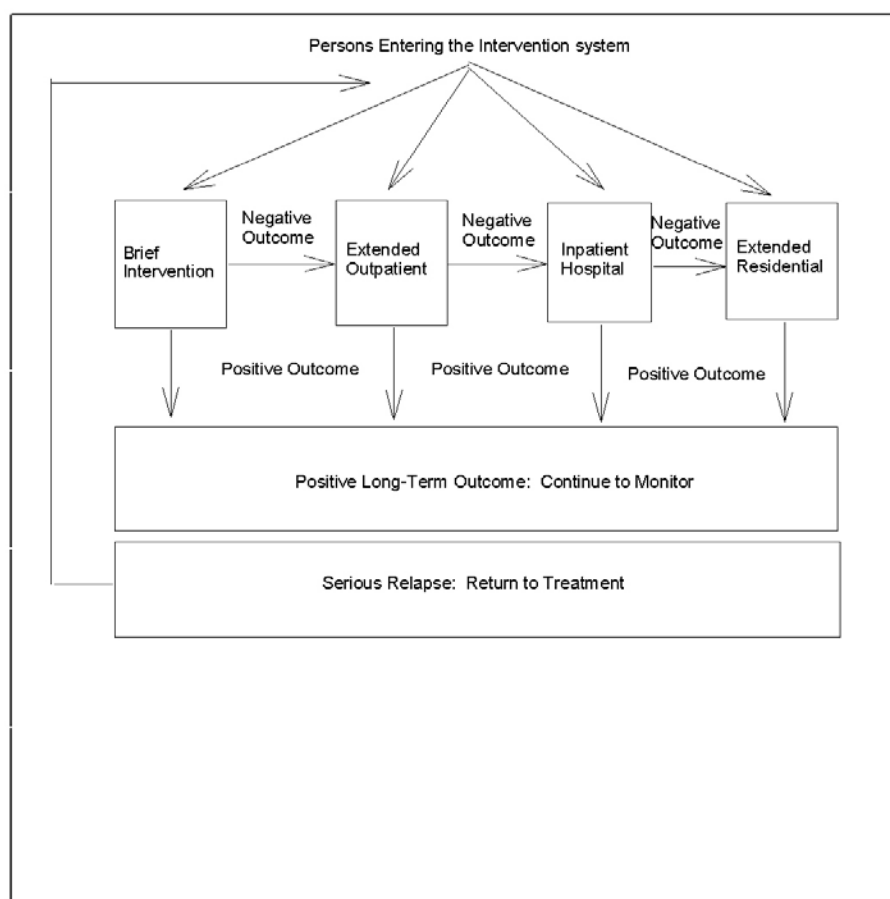
Fortunately, we do not feel much pressure to campaign strongly for neuropharmacological and supporting basic neurochemical research on AG, in the way we do where BE and NADG are concerned. The relevant research has lately exploded and will clearly continue to do so. Among the reasons for this is that a patentable, effective pharmacological instrument for treating AG would very handsomely reward its developers. This in itself might raise a cautionary note where South Africa is concerned, given our history of difficult political and legal issues around access to patented medicines at prices affordable in a middle-income country with many very poor people. If there is lobbying that is in order where AG pharmacological research is concerned, it is plausibly to be done around making sure that AG is regarded fully seriously as the life-threatening and *real* disorder it is, so that future drugs for treating AG are brought under the umbrella of more liberal regimes for developing-country access to essential medicines under negotiation in the World Trade Organization, and are not classified as luxury or lifestyle medicines.

Finally, though we are sanguine about neuropharmacological research on AG continuing to be well supported, we reiterate a point we made in section 4.3 concerning the need for *integrated* research that includes BE. We have seen that the recent breakthroughs in understanding addiction in general and AG in particular rested on the conjunction of basic neurochemistry and neuroeconomics. In section 4.3 we noted that neuroeconomics relies heavily on fMRI experiments which in turn require batteries of well-tested behavioral task protocols from BE. Let us give an example. In Chapter 3 we reported on an experiment by Nieuwenhuis and colleagues³³⁷ that explored effects of available outcome contexts on reward system response. With slight adaptations, this protocol could be adapted to directly test Rachlin's mechanism for disastrous win-loss accounting by PGs. Such an experiment might never occur to scientists mainly motivated by a neuropharmacological perspective; yet we think it would shed light on some open questions about DG that have come up repeatedly in this survey.

At present there is little capacity for basic (as opposed to clinical) fMRI research in South Africa. We hope that this situation will soon improve, since in its absence we cannot expect to be able to train basic neuroscientists. However, there is capacity for supporting BE work of the kind we have identified as part of the next wave of research that, we hope, will soon turn AG from a serious social problem into a minor nuisance.

³³⁷ Nieuwenhuis, S., Heslenfeld, D., Alting von Geusau, N.J., Mars, R., Holroyd, C., and Yeung, N. (2004). Activity in human reward-sensitive brain areas is strongly context dependent. *NeuroImage* 25: 1302-1309.

Figure 1: A Stepped Care Intervention System



We turn now to specific personal-level intervention recommendations based on BE. Earlier we alluded to the fact that BE, as a conceptual framework, has not been very influential in the design of interventions for alcohol and drug problems or for DG. We wish to emphasize that this relative lack of influence *is not* because BE lacks a general and sophisticated theoretical framework and a substantial empirical data base, which we reviewed in Chapter 1. Instead, the relative neglect of BE among substance abuse treatment researchers has been due to the non-behavioral (and sometimes even anti-behavioral) intellectual commitments among leaders in the treatment field³³⁸. This appears to be changing. Indeed, the alcohol treatment community has amply shown that it is sensitive to scientific developments, and presumably that will be true for the DG community as well.

³³⁸ Vuchinich, R.E. (1995). Alcohol abuse as molar choice: An update of a 1982 proposal. *Psychology of Addictive Behaviors* 9: 223-235.

Just two or three decades ago, the ideas that now dominate alcohol treatment research (i.e., cognitive-social learning theory³³⁹, stages of change³⁴⁰, relapse prevention³⁴¹, and motivational enhancement³⁴²) were not yet on the scene or were only minority approaches. They were brought to prominence by persuasive empirical evidence, and we can think of no reason why this cannot happen with BE, provided, of course, that research from this perspective continues to find supportive evidence. BE has extensive basic science support and a rapidly growing applied research base in the alcohol and drug abuse fields, and this is now being extended to the NADG field.

Recommendations from Ainslie's model

Since reward bundling (see Figure 6) is the critical self-control strategy in Ainslie's model, the most general recommendation is to develop ways to promote reward bundling among NADGs. Although the theory covering reward bundling is extensive (e.g.,³⁴³), and we know that both animals³⁴⁴ and people³⁴⁵ will bundle future rewards under carefully constructed laboratory conditions, little scientific evidence currently is available regarding how reward bundling develops and changes in humans in their natural environments. So the following are theory-based speculations, but each would be an important topic for future research.

Encouraging reward bundling almost certainly would entail changes in the way that NADGs perceive their choice situation. As Rachlin³⁴⁶ has noted, "Perhaps the most serious obstacle to self-control is faulty perception of one's own behavior. It is much more difficult for an alcoholic to perceive how much he is drinking than it is for his friends and family members to perceive how much he is drinking." There are several ways in which increased bundling of rewards and perceptual changes could be accomplished. First, having NADGs keep a record (self-monitoring) of their gambling behavior could produce positive results, as has been found in some alcohol treatment studies³⁴⁷.

-
- ³³⁹ Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice-Hall.
- ³⁴⁰ Prochaska, J.O., and DiClemente, C. (1983). Stages and processes of self-change in smoking: Toward an integrative model. *Journal of Consulting and Clinical Psychology* 51: 390-395.
- ³⁴¹ Marlatt, G.A., and Gordon, J. (1985). *Relapse prevention*. New York: Guilford.
- ³⁴² Miller, W.R., Rollnick, S., & Conforti, K. (2002). *Motivational interviewing: Preparing people for change* (2nd Edition). New York: Guilford.
- ³⁴³ Ainslie, G. (1992). *Picoeconomics: The strategic interaction of successive motivational states within the person*. Cambridge, England: Cambridge University Press.
- ³⁴⁴ Ainslie, G., and Monterosso, J. (2003). Building blocks of self-control: Increased tolerance for delay with bundled rewards. *Journal of the Experimental Analysis of behavior* 48: 37-48.
- ³⁴⁵ Kirby, K.N., and Guastello, B. (2001). Making choices in anticipation of similar future choices can increase self-control. *Journal of Experimental Psychology: Applied* 7: 154-164.
- ³⁴⁶ Rachlin, H. (2000). *The science of self-control*. Cambridge, MA: Harvard University Press.
- ³⁴⁷ Clifford, P.R., and Maisto, S.A. (2000). Subject reactivity effects and alcohol treatment outcome. *Journal of Studies on Alcohol* 61: 787-793.

Rachlin³⁴⁸ reports a study in which smokers were asked to self-monitor their smoking but were not asked to decrease or increase their smoking. Perhaps surprisingly, the relatively simple act of self-monitoring produced reductions in smoking in many subjects. The BE theoretical explanation of this therapeutic effect is that the act of monitoring strongly encourages the individual to focus on the longer term pattern of their actions, rather than just on the current acts. Rachlin explains: “Just smoking a cigarette that *happens* to be the twenty-fifth cigarette smoked that day is quite different from smoking a cigarette that is *known* to be the twenty-fifth cigarette smoked that day. In the former case, the act of smoking the cigarette stands alone as the most valuable alternative available within the next five minutes; in the latter case, smoking this cigarette is part of a larger pattern that is much less highly valued.” Developing easily used self-monitoring systems for NADGs would be a relatively simple undertaking, and, if effective, would be easily implemented in public health and clinical settings.

Second, reward bundling among NADGs could be facilitated by helping them formulate personal rules around gambling. For example, a NADG could make a rule to “gamble only on dedicated casino holidays,” or “only on a particular day of the week or month,” or “gamble only with overtime pay,” etc. Of course, such personal rules would have to be extensively tailored to the individual NADG, based on her life circumstances, belief systems, and cognitive capacity. Also, because personal rules are subject to being undermined by coalitions of interests, the role of friends, family, and (perhaps) counselors would be particularly important in buttressing this sort of strategy. This is the case because the NADG will find it less easy to create new exceptions to personal rules if the interpretation of these rules is a matter for interpersonal negotiation and is not strictly up to him or her. This also suggests a policy intervention: government and/or industry might subsidize programs that would educate the families and friends of problem gamblers on BE theory of bundling and personal-rule formation and maintenance.

A third option for promoting reward bundling is to borrow the motivational interviewing (MI) intervention now rapidly spreading through the substance abuse and addiction field³⁴⁹. The theory, empirical evidence for, and modes of implementation of MI are largely due to the work of psychologist William R. Miller.³⁵⁰ Miller noted that, given their commitment to the disease model, most traditional addiction treatments were aggressive and confrontational, in order to “break down denial,” and were very directive regarding the necessary steps to recovery. Miller’s research showed, however, that clients treated by empathic, non-directive counselors had much better outcomes 12 months later than did clients treated by confrontational, directive counselors. So Miller developed MI as a non-confrontational, relatively non-directive technique, the main goal of which is to increase the client’s motivation to change their behavior. In MI, the client is engaged in a non-judgmental conversation about the plusses and minuses of their drinking (or smoking or etc.), especially insofar as it impacts their long-term goals. One primary aim of MI is to develop *discrepancy* between a person’s likely success in reaching their long-term goals if they continue their substance use, on the one hand, and their likely success with long-term goals if they curtail or stop their substance use, on the other hand. Clearly perceiving such discrepancies, according to MI theory, increases the clients’ motivation to change.

³⁴⁸ Rachlin, H. (2000). *The science of self-control*. Cambridge, MA: Harvard University Press.

³⁴⁹ Miller, W.R., Rollnick, S., and Conforti, K. (2002). *Motivational interviewing: Preparing people for change* (2nd Edition). New York: Guilford.

³⁵⁰ *Ibid.*

Although the processes through which MI exerts its effects are not well understood³⁵¹, such changes in motivation are fairly easily understood within the BE framework. It is quite possible the MI counseling is efficacious because it gets clients to bundle rewards more than previously, in which case it should also may be helpful for NADGs.

Recommendations from Rachlin's model

If Rachlin is on the right track with his relative addiction theory (see Chapter 1, section 6), then measures should be sought that increase the price of entry to gambling situations for NADGs, and/or decrease their price of entry to situations of social interaction. Recall, however, that we have argued that NADGs, in contrast to AGs, cannot be picked out by distinctive personal properties. Thus we need to consider measures for changing relative costs that don't require this. One such possibility is imposing cover charges on entry to casinos that vary inversely with the size of parties.³⁵² Of course, lone NADGs could try to get around such measures by trying to nominally attach themselves to other parties at the casino entrance; but even this would require that they devote some minimum portion of their investment to social interaction. Casinos could also be required to maintain facilities that emphasize social activities other than gambling, and that thus encourage family members to accompany gamblers on their outings. The Grand West casino in Cape Town, for example, has extensive shopping, restaurant, day-care and even museum facilities that surround the gambling floor and are fully as visually salient as it is.

Rachlin's model helps to explain a fact that is increasingly acknowledged in the gambling policy, regulation and epidemiology literature: slot machine gambling is a more serious threat for NADGs than casino or racetrack gambling, and internet gambling may be the most dangerous of all. At a poker table, roulette wheel or racetrack, the gambler at least has a minimum level of interaction with fellow gamblers. In the detached module of the slot machine interface, by contrast, and even more so in the complete privacy of his or her home internet port, the gambler can withdraw completely, and thereby, over time, increase the cost of re-entry to social circumstances. This point should not be taken to suggest, however, that all social interaction, especially interaction with other gamblers, is necessarily a good thing according to Rachlin's model. Some specific forms of social interaction can become *complements* to, rather than substitutes for, gambling; that is, they can be goods which, when consumed, encourage higher consumption of gambling. (Consider, by analogy, cold meats and mustard, or MP3 players and downloaded music, or, among the addictions, alcohol and cigarettes.) Rather, the point is simply that forms of gambling that lead the gambler to isolate himself are apt to be unusually addictive if Rachlin's model is correct; and evidence suggests that indeed they are. Should this imply more stringent regulation of slot machines and internet gambling than of multi-person table games and sports betting?

Rachlin's model adds at least one further suggestion to the list of personal policies discussed in connection with the more generic BE account. This is that gamblers should try to confine their activity to familiar venues where they know people, rather than raise the cost of social interaction (thus lowering the relative cost of increased gambling) by putting themselves in unfamiliar betting environments where they are strangers.

³⁵¹ Miller, W.R. (1996). Motivational interviewing: Research, practice, puzzles. *Addictive Behaviors* 21: 835-842.

³⁵² In case this sounds novel to the point of being bizarre, note that it has precedents in some nightclubs that admit men, or reduce their entry fee, only if they are accompanied by women.

Finally, successfully addressing the perceptual distortions among NADGs that Rachlin proposed in his specific theory of DG discussed in Chapter 1, section 7³⁵³, potentially could have large therapeutic effects for NADGs. Recall that different individuals will develop different levels of aggregations of bets over which to do their utility bookkeeping. An intervention that modified NADGs' aggregation of betting units would be quite useful. This has never, to our knowledge, *directly* been attempted and how difficult it would be is not now known³⁵⁴. However, what is reasonably clear is that some standard casino policies and practices are effectively designed to interfere with cognitive record-keeping in general, including both record-keeping relevant to the Rachlin mechanism and record-keeping relevant to bundling. In the next section, we will discuss some research that we recommend to test the impact of casino environments on cognitive record-keeping, and possible regulatory changes that could reduce prevalence of NADG to which they contribute.

³⁵³ Rachlin, H. (1990). Why do people gamble and keep gambling despite heavy losses? *Psychological Science* 1: 294-297.

³⁵⁴ Delfabbro, P. (2004). The stubborn logic of regular gamblers: obstacles and dilemmas in cognitive gambling research. *Journal of Gambling Studies* 20: 1-21.

5. Appendix: Disordered gambling and gender

Summary

There are important differences between males and females with reference to disordered gambling. Males account for a significantly greater number of disordered gamblers than females. Males tend to start gambling younger than females, and take longer than females to become disordered gamblers. Males and females tend to prefer different types of gambling, and differ in the sorts of problems associated with their disordered gambling. These differences should inform the direction of future research, as well as measures intended to identify people at risk of disordered gambling, and to deal with treatment of those who have become disordered gamblers.

In many ways, of course, all people irrespective of gender are very much the same as one another. Male and female problem gamblers are also similar to one another in a wide range of important respects, including their physiological responses to gambling,³⁵⁵ and in having greater rates of substance abuse and suicide³⁵⁶ than those who do not pay to participate in games of chance at all, and non-disordered gamblers. Nonetheless differences in the processes of physical development for males and females, and hormonal and other differences, including often sustained differences in treatment and expectations, over the course of life, are associated with measurable differences from the point of view of the sciences of decision, including differences in prevalence of different types of gambling activity and of disordered gambling between males and females.

These differences should inform programmes oriented towards the prevention and treatment of disordered gambling, and the design of future research aimed at better understanding of disordered gambling. Here is a brief review of some of the more recent findings.

Prevalence of gambling and disordered gambling

There seems to be relatively little difference in the prevalence of regular or non-disordered gambling for males and females. The 1998 Gambling Impact and Behavior Study in the United States found a lifetime gambling rate of 88% for males and 83% for females.³⁵⁷

³⁵⁵ See, for example, Coventry, K. R. and Hudson, J. (2001). Gender differences, physiological arousal and the role of winning in fruit machine gamblers. *Addiction*, 96, 871–879.

³⁵⁶ Volberg, R. A. (1994). The prevalence and demographics of pathological gamblers: Implications for public health. *American Journal of Public Health*, 84, 237–241.

³⁵⁷ Volberg, R., Harwood, H., & Tucker, A. (1999). Gambling impact and behavior study. National Opinion Research Center at the University of Chicago, Chicago, IL.

Similar results have been found in other studies,³⁵⁸ although different ways of assessing prevalence³⁵⁹ need not lead to the same relative prevalence for males and females being reported. One recent study found lower frequency of gambling in homosexual males than heterosexuals, and a corresponding higher frequency in homosexual females compared to heterosexuals.³⁶⁰ This may seem surprising, but the result was what the authors of the study had hypothesised, given the view that sexual orientation is to a significant extent determined by hormonal processes taking place during fetal development.³⁶¹ Nonetheless few prevalence studies report significant differences. The National Responsible Gambling Programme in South Africa reported in 2003 that regular gamblers (lottery only) were very evenly split between males and females, while females accounted for a larger share of the occasional and never gambled groups.³⁶²

There are, though, clear differences in the prevalence of disordered gambling. More specifically, males account for a disproportionately large number of disordered gamblers given the similar prevalence rates.^{363,364,365}

-
- ³⁵⁸ See, for example, Welte, J. W., Barnes, G. M., Wieczorek, W., Tidwell, M. C., & Parker, J. (2002). Gambling participation in the US results from a national survey. *Journal of Gambling Studies*, 18, 313–337.
- ³⁵⁹ For example, measuring frequency of gambling activity, rather than distinguishing people who gambled at all from those who never pay to participate in games of chance, or place a bet in a conventionally recognised gambling setting, need not lead to the same relative prevalence for males and females.
- ³⁶⁰ Hershberger, S. L. and Bogaert, A. F. (2005). Male and female sexual orientation differences in gambling. *Personality and Individual Differences*, 38, 1401–1411.
- ³⁶¹ Ellis, L., & Ames, M. A. (1987). Neurohormonal functioning and sexual orientation: a theory of homosexuality–heterosexuality. *Psychological Bulletin*, 101, 233–258.
- ³⁶² National Responsible Gambling Programme (2003) Gambling and Problem Gambling in South Africa: A National Study 2003.
- ³⁶³ Hraba, J., & Lee, G. (1996). Gender, gambling and problem gambling. *Journal of Gambling Studies*, 12, 83–101. Ladd, G. T., & Petry, N. M. (2002). Gender differences among pathological gamblers seeking treatment. *Experimental and Clinical Psychopharmacology*, 10, 302–309.
- ³⁶⁴ Ladd, G. T., & Petry, N. M. (2002). Gender differences among pathological gamblers seeking treatment. *Experimental and Clinical Psychopharmacology*, 10, 302–309.
- ³⁶⁵ Martins, S. S., Lobo, D. S. S., T. W., Barnes, G. M., Wieczorek, W., Tidwell, M. C., & Parker, J. (2002). Gambling participation in the US results from a national survey. *Journal of Gambling Studies*, 18, 313–337.
- ³⁶⁵ For example, measuring frequency of gambling activity, rather than distinguishing people who gambled at all from those who never pay to participate in games of chance, or place a bet in a conventionally recognised gambling setting, need not lead to the same relative prevalence for males and females.
- ³⁶⁵ Hershberger, S. L. and Bogaert, A. F. (2005). Male and female sexual orientation differences in gambling. *Personality and Individual Differences*, 38, 1401–1411.
- ³⁶⁵ Ellis, L., & Ames, M. A. (1987). Neurohormonal functioning and sexual orientation: a theory of homosexuality–heterosexuality. *Psychological Bulletin*, 101, 233–258.

A 1999 study in the United States found that males accounted for nearly two thirds of the disordered gamblers in the 18 states studied.³⁶⁶ The same proportion of male to female disordered gamblers was found in a study at around the same time in Canada.³⁶⁷ In Switzerland it has been found that rates of problem gambling are around three times higher in males than in females.³⁶⁸ One comparative study of seven countries found relative over-representation of males, especially young males, in the number of disordered gamblers.³⁶⁹ A meta-analysis of research on prevalence of disordered gambling in the United States and Canada found that males formed a greater portion of disordered gamblers than females whether the data was first partitioned according to status as an adolescent, a college student, member of a treatment programme or prisoner.³⁷⁰ We are not aware of recent figures on the relative proportion of disordered gamblers that are male in South Africa. This issue is not discussed in the 2003 report we refer to a few paragraphs above.

Further differences

There are also clear indications of differences in the sorts of gambling activity preferred by males and females, and the ways in which disordered gambling manifests, and in how maladaptive decision-making can be cued.

As some put it, male disordered gamblers are more likely to be 'action gamblers' who begin gambling younger,³⁷¹ are disordered gamblers for longer, and focus on 'skill' games such as track betting or blackjack. Female disordered gamblers on the other hand tend to be 'escape' gamblers who become problem gamblers more quickly, and favour 'luck' games such as slot machines and bingo.³⁷²

-
- ³⁶⁵ National Responsible Gambling Programme avares, H. and Gentil, v. (2002). Pathological Gambling In Women: A Review. *Rev. Hosp. Clín. Fac. Med. S. Paulo* 57(5), 235-242.
- ³⁶⁶ The median proportion of women among disordered gamblers was 38%, with the remainder (62%) males. National Research Council. (1999). *Pathological gambling: A critical review*. The National Academies Press, Washington, DC.
- ³⁶⁷ More specifically, the same result was found in Alberta. In other provinces the fraction of disordered gamblers that was male was ever greater. National Council of Welfare. (1996). *Gambling in Canada* (Catalogue No. H68-40/1996E). Minister of Supply and Services Canada, Ottawa, ON.
- ³⁶⁸ Bondolfi, G., Osiek, C., & Ferrero, F. (2000) Prevalence estimates of pathological gambling in Switzerland. *Acta Psychiatrica Scandinavica*, 60, 601-609.
- ³⁶⁹ Gray, P. B. (2004). Evolutionary and Cross-Cultural Perspectives on Gambling. *Journal of Gambling Studies*, 20(4), 347-371.
- ³⁷⁰ Shaffer, H. J., Hall, M. N., & Bilt, J. V. (1997). Estimating the prevalence of disordered gambling behavior in the United States and Canada: A meta-analysis. *Harvard Medical School Division on Addictions*, Harvard University.
- ³⁷¹ This result has also been found in New Zealand: Abbott, M. W. (2001). Problem and non-problem gamblers in New Zealand: A report on phase two of the 1999 National Prevalence Survey. Report number six of the New Zealand Gaming Survey. The Department of Internal Affairs, Wellington, New Zealand.
- ³⁷² Lesieur, H. R., & Blume, S. B. (1991). When lady luck loses: Women and compulsive gambling. In N. Van Den Bergh (Ed.), *Feminist perspectives on addictions* (pp. 181-197). New York: Springer. See also Gray, P. B. (2004). Evolutionary and Cross-Cultural Perspectives on Gambling. *Journal of Gambling Studies*, 20(4), 347-371.

Other research makes the same general point, but expresses the distinction as being one between 'strategic' and 'non-strategic' forms of gambling, noting that male gamblers were more likely to report problems with the former, females with the latter.³⁷³ The same study distinguishing strategic and non-strategic gambling found that male disordered gamblers were more likely than females to report an arrest related to gambling, or a drug problem, whereas females were more likely than male disordered gamblers to report having received mental health treatment (not related to disordered gambling). This same study found that female gamblers were more likely to report anxiety and suicide attempts attributed to gambling. Male gamblers' debts were more likely to be to loan sharks and bookies, while for women debt was more likely through formal credit facilities. Finally, this study confirmed the finding that the route from beginning gambling to treatment seeking disordered gambler was faster for females than males.³⁷⁴ Female disordered gamblers, furthermore, appear to have a greater suicide rate than males, whereas males are more likely than females to engage in risky sexual behaviour and to abuse alcohol.³⁷⁵ Other work suggests that males generally have a more positive attitude to gambling.³⁷⁶

One study assessing temporal discounting behaviour (see chapter 2-1) in males and females found that negative affect in women lead to increased impulsiveness, whereas the same effect in males followed positive affect.³⁷⁷ This result should be read alongside the finding that induction of anger led to male subjects being less risk averse (taking more risks and risks of greater size), while induction of disgust led to female subjects being more risk averse (taking fewer risks, and risks of smaller size).

-
- ³⁷³ Ladd, G. T., & Petry, N. M. (2002). Gender differences among pathological gamblers seeking treatment. *Experimental and Clinical Psychopharmacology*, 10, 302–309. This paper reports a female preference for slot machines over gambling on sports and card games. See also McDaniel, S. R., and Zuckerman, M. (2003). The relationship of impulsive sensation seeking and gender to interest and participation in gambling activities. *Personality and Individual Differences*, 35, 1385–1400.
- ³⁷⁴ Potenza, M. N., Steinberg, M. A., McLaughlin, S. D., Wu, R., Rounsaville, B. J., and O'Malley, S. S. (2001). Gender-Related Differences in the Characteristics of Problem Gamblers Using a Gambling Helpline. *American Journal of Psychiatry*, 158, 1500–1505. See also Hraba, J., & Lee, G. (1996). Gender, gambling and problem gambling. *Journal of Gambling Studies*, 12, 83–101.
- ³⁷⁵ Martins, S. S., Tavares, H., Da Silva Lobo, D. S., Galettia, A. M., Gentil, V. (2004). Pathological gambling, gender, and risk-taking behaviors. *Addictive Behaviors*, 29, 1231–1235.
- ³⁷⁶ Skitch, S. A. and Hodgins, D. C. (2005). A Passion for the Game: Problem Gambling and Passion Among University Students. *Canadian Journal of Behavioural Science*, 37(3), 193–197. Kassinove, J. I. (1998). Development of the gambling attitude gambling scales: Preliminary findings. *Journal of Clinical Psychology*, 54, 763–771.
- ³⁷⁷ Oxoby, R. J and McLeish, K. N. (2004). Gender, Affect and Intertemporal Consistency: An Experimental Approach. University of Calgary, Department of Economics Discussion Paper 2004-17.

Conclusion

Future research into disordered gambling should be sensitive to the research briefly described here. Subjects pools should as far as possible be large enough to have a prospect of significant results for male and female participants. Experiments and attempts to study manipulations of decision making in order better to understand the onset of disordered gambling or its treatment should be alert to the differences described here, and the possibility of others like them.