RISK PREFERENCES, TIME PREFERENCES AND SMOKING BEHAVIOUR

by

Glenn W. Harrison, Andre Hofmeyr, Don Ross and J. Todd Swarthout†

February 2018

Forthcoming, *Southern Economic Journal*

ABSTRACT

There is a rich theoretical literature in economics which models habit-forming behaviours, of which addiction is the exemplar, but there is a paucity of experimental economic studies eliciting and comparing the preferences that economic theory suggests may differ between addicts and non-addicts. We evaluate an incentive-compatible risk and time preference experiment conducted on a sample of student smokers and non-smokers at the University of Cape Town in 2012. We adopt a full information maximum likelihood statistical framework, which is consistent with the data generating processes proposed by structural theories and accounts for subject errors in decision making, to explore the relationship between risk preferences, time preferences and addiction. Across different theories and econometric specifications we find no differences in the risk preferences of smokers and non-smokers but do find that smokers discount significantly more heavily than non-smokers. We also identify a nonlinear effect of smoking intensity on discounting behaviour and find that smoking intensity increases the likelihood of discounting hyperbolically, which means heavier smokers may be more prone to time inconsistency and more recalcitrant to treatment. These results highlight the importance of the theory-experimental design-econometric trinity and have important implications for theories of addiction.

Keywords: smoking, discount rates, risk aversion, time inconsistency, addiction

JEL codes: I1, D81, D91

† Department of Risk Management & Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA (Harrison); School of Economics, University of Cape Town, South Africa (Hofmeyr); Department of Philosophy, University College Cork, Ireland, School of Economics, University of Cape Town, South Africa, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA (Ross); and Department of Economics, Andrew Young School of Policy Studies, Georgia State University, USA (Swarthout). Harrison is also affiliated with the School of Economics, University of Cape Town. E-mail contacts: gharrison@gsu.edu, andre.hofmeyr@uct.ac.za, don.ross931@gmail.com and swarthout@gsu.edu. We are grateful to the University of Cape Town and the American Cancer Society for funding this research. These funders had no involvement in any aspect of the research project nor in the preparation of this manuscript.

I. INTRODUCTION

Addiction is a puzzle for economic theory: how can rational-agent modelling accommodate the fact that most addicts expend resources to acquire their targets of addiction but simultaneously incur real costs to try to reduce or limit their consumption of these goods? Furthermore, why is the typical course of addiction characterised by repeated unsuccessful attempts to quit prior to final abstention? From the standpoint of standard consumer theory in economics these patterns of behaviour are difficult to rationalise.

A number of economists over the years have risen to the challenge. In Section II we review these efforts, and conclude that making further progress, especially in critically bringing economic modelling of addiction to bear on psychological and clinical studies, requires as a first step more rigorous specification and identification of the relationships between structural risk and time preferences, on the one hand, and statistical vulnerability to addiction, on the other. Such progress requires careful experimentation to calibrate parametric relationships among preference structures and choices that generate, sustain, and mitigate addiction. We undertake such experimentation, using regular smokers as the representative addicts.

An incentive-compatible experimental design allows us to explore potential differences in the risk and time preferences of smokers and non-smokers and jointly estimate utility function curvature and discounting functions. We find no significant differences in the risk preferences of smokers and non-smokers but do *find that smokers discount the future significantly more heavily than non-smokers*. These results are robust to different assumptions about the way people evaluate lotteries and the way they discount utility flows. In addition, we identify *a nonlinear effect of smoking intensity on discounting behaviour* and find that *smoking intensity increases the likelihood of discounting hyperbolically*, which, under the assumption of an additively-separable intertemporal utility function, means smokers, and in particular, heavier smokers, may be more prone to time inconsistency.

This research makes a number of contributions to the literature. Instead of adopting the standard two-step approach to data analysis (see Section III), which is statistically invalid, we estimate risk and time preference parameters as a linear function of observable characteristics (e.g., age, gender, and smoking status) so that the uncertainty of the parameter estimates propagates into the inferences which are drawn from the data.

In addition, when analysing risk preferences and smoking behaviour, we allow risk attitudes to be determined both by utility function curvature and probability weighting. Prior studies in the literature either focus on utility function curvature or probability weighting, but not on both together.

This is only the second study in the smoking-discounting literature to incorporate utility function curvature in the estimation of time preference models, and it is the first which allows rank-dependent utility theory to characterise choice under risk. In addition, this is the first study to identify a nonlinear relationship between smoking intensity and discounting behaviour. Smoking more cigarettes is associated with increased discounting but only up to a point, after which each additional cigarette is associated with lower discounting.

The design and analysis are sensitive to the recognition that multiple decision processes characterise the discounting of delayed rewards. It is crucial for researchers to be cognisant of this fact when exploring the addiction-discounting relationship. Smoking intensity increases the likelihood of discounting hyperbolically, which may be an important factor in tobacco addiction and explain recalcitrance to treatment.

Following the review of economic models of addiction in Section II, Section III reviews previous research on the relationship between risk preferences, time preferences and smoking behaviour. Section IV discusses our experimental design and presents summary statistics for the sample. Section V formulates our statistical approach to data analysis. Section VI presents the results and Section VII concludes.

II. ECONOMIC MODELS OF ADDICTION

Existing work by economists in modelling addictive consumption may be grouped into two broad approaches.[1]

The first approach, often referred to in the literature as *rational addict* modelling, was pioneered by Becker and Murphy (1988). It attributes addiction to unusual properties of certain goods, which causes flows of utility from their consumption to accumulate as capital that incentivises further consumption and reduces marginal utility from non-addictive substitutes. On this kind of account, agents fall into addiction without at any point behaving contrary to their consistent preferences, and it is not even necessary to posit uncertainty about outcomes or forecasts of utility.

Rational addiction models have been widely criticised for systematically mis-predicting the patterns of temporary cessation and relapse, followed by eventual success in achieving control, that characterises the typical life course of an addiction (e.g., Ross (2010)). The natural prediction of the basic Becker and Murphy (1988) model is that an addict will simply keep consuming their addictive target unless and until its price rises beyond the point where its consumption is optimal at the moment of choice. The model does, however, offer a prediction, which psychologists have generally considered reasonable, about the characteristics of people who are likely to be most vulnerable to addiction: those who discount future utility most steeply.

Orphanides and Zervos (1995) added an additional dimension to rational addict modelling by incorporating uncertainty on the part of potential consumers about the extent of their vulnerability to addiction when they first sample potentially addictive goods. This model yields the further prediction, which has again been regarded by

[1] Outside of the two general approaches we review, some economists have favoured models in which the dynamics of addictive processes occur outside the logical space of economic agency, even if within the brain and nervous system of the person (e.g., Laibson (2001), Loewenstein, O'Donoghue and Rabin (2003), Gul and Pesendorfer (2007)). In such models, addictive temptations are exogenous sources of costs to maintenance of consistent or welfare-maximising choice that under some circumstances overwhelm the agent's budget of resources for resistance. For further discussion of these models see Ross (2011, 2014a, 2014b).

many addiction scientists and clinicians as intuitive, that risk aversion, both instantaneous[2] and intertemporal, should be a protective factor against addiction.

The second broad approach to economic modelling of addiction responds to criticisms of rational addict models for failing to capture the observed synchronic and diachronic preference ambivalence of most addicts that is reflected in their apparent efforts to resist and modify their own revealed preferences for addictive goods. Economists have attempted to deal with this by complicating the agency of addicts in one or both of two ways: with either diachronic or synchronic dual self models.

Diachonic dual self models (Winston (1980), Thaler and Shefrin (1981), Schelling (1984), Gruber and Köszegi (2001), Bénabou and Tirole (2004)) divide the addicted agent into temporal successions of sub-agents that implement divergent temporal discounting functions. Both Gruber and Köszegi (2001) and Bénabou and Tirole (2004) incorporate the quasi-hyperbolic intertemporal discounting model of Laibson (1997) to explain why addicts choose to consume addictive targets at a present moment while simultaneously preferring to refrain from such consumption in the future. Such a pattern implies inconsistent choice over time by the succession of sub-agents considered as a group. Diachronic dual self models can then capture varying levels of success in resolving such ambivalence by allowing for variation in the extent to which addicts accurately recall or predict their own preference histories and courses. Consequently, these models also often involve choice under uncertainty.

By contrast, *synchronic dual self* models incorporate sub-agents that compete for control of the agent's choices at a given point in time (Benhabib and Bisin (2004), Bernheim and Rangel (2004)). In these models, the competing agents again differ from one another in the intertemporal discounting behaviour that they implement when they respectively gain control, and also face varying degrees of uncertainty concerning the implications of addictive consumption for present welfare, future welfare, or both. Fudenberg and Levine (2006, 2011, 2012) develop models that

[2] The prefix "instantaneous" is used to differentiate instantaneous risk preferences from intertemporal risk preferences. Intertemporal risk preferences refer to preferences over intertemporal lotteries, the outcomes of which may be temporally correlated. By contrast, instantaneous risk preferences define atemporal attitudes to risk and uncertainty. We only empirically examine instantaneous risk preferences so all subsequent references to "risk preferences" refer to the instantaneous or atemporal variety.

combine diachronic and synchronic complexity of agency. While varying in their details and the specific behavioural phenomena they are designed to capture, the three Fudenberg and Levine models share as their core strategic interaction and partial conflict between short-run sub-agents ("selves") that are relatively less patient than, and relatively more risk averse than, long-run sub-agents ("selves").

As we document in Section III with specific reference to addictive smoking, psychological studies of addiction have also focused recurrently on steep temporal discounting and relative indifference to risk as factors that may contribute to the formation and persistence of addiction; for a review of psychological literature of this kind going beyond smoking, see Ross et al. (2008) chapters 3 and 4. There is, furthermore, increasing consensus among psychologists that addictions are learned, and modifiable by incentivising interventions (Redish, Jensen and Johnson (2008), Heyman (2009)). Psychologists might therefore be expected to welcome efforts by economists to contribute improved specification precision and technical rigour with respect to the empirical identification of risk and time preference idiosyncrasies that distinguish addicts.

It thus constitutes a significant gap in the literature that economists have not yet directly empirically estimated differences in risk and time preferences, specified with full theoretical precision, between addicts and non-addicts. An important aspect of such precision is to respect the need for joint estimation of risk and time preferences established by Andersen et al. (2008). Unsurprisingly, none of the many empirical studies by psychologists of temporal discounting differences between addicts and non-addicts attempt, or indeed recognise the importance of, such joint estimation. In its absence, as Andersen et al. (2008) demonstrate theoretically and empirically, discount rate estimates are significantly biased upward for risk averse agents, which is also likely to result in mis-estimation of whether their structure is exponential, hyperbolic, or quasi-hyperbolic. Nor have structural interactions between risk and time preferences been explicitly specified in existing economic models of addiction. Such specification as it might feature in the distinguishing characterisation of addicts cannot be based on a priori theorising, but depends on empirical data.

Our empirical comparison of temporally indexed and risky choice behaviour in a sample of smokers and a sample of non-smokers is motivated by this concern with improved economic modelling of addiction in general. We chose to study smokers for three reasons: nicotine is the most readily available addictive drug in general populations; there is widespread agreement among addiction scientists and clinicians that almost all regular, daily smokers meet the criteria for addiction (West (2006)); and the relative non-interference of nicotine with basic cognition and judgment makes nicotine addicts a natural starting point population for any new laboratory paradigm.

In our view, improved unification of economic and psychological approaches to addiction is most likely to the extent that research in both disciplines is alert to a self-conscious philosophical orientation. We are guided by the approach that Ainslie (1992, 2001) has dubbed "picoeconomics" (see also Ross et al. (2008)). This approach emphasises, as does Heyman (2009), that addiction is in large part learned behaviour, expressed through choices that are 'voluntary' in the non-metaphysical sense of being responsive to incentives.

The framework of Ainslie (1992, 2001) recognises that both exogenous and endogenous neurophysiological and neurochemical states and processes give rise to vulnerabilities and barriers to controlling addiction that an economic model will represent as variable costs. The picoeconomic model emphasises the role of inconsistent intertemporal discounting in generating and maintaining addictive choice patterns, but it does *not* predict, counterfactually, that most human choice over time reflects hyperbolic discounting. Rather, it applies a philosophical thesis that consistent valuation of rewards over time requires explanation and should not simply be assumed as a natural default disposition. Ainslie himself emphasises the importance of "personal rules," that is, self-enforcing linkages between discrete choices that should be reflected in agents' revealed preferences, but he is also alert to the importance of institutional and other environmental "scaffolding" (Clark (1997)) as providing support for intertemporally consistent valuation and choice.

Economists and psychologists, notwithstanding their different practical priorities, can join in seeking explanation of addiction in breakdowns and loopholes in personal rules, in challenges to their implementation resulting from errors in risk perception and estimation, and in strategic complications in the relationships between individuals and their social environments.

III. REVIEW OF THE LITERATURE ON RISK PREFERENCES, TIME PREFERENCES AND SMOKING BEHAVIOUR

Smoking is known to be one of the primary behavioural risk factors for the additional utilisation of health resources and expenditures on health. For just over 50 years the U.S. Surgeon General has been collating careful epidemiological evaluations of the causal effect of smoking on a large number of diseases (U.S. Department of Health and Human Services (2014)). And major litigation efforts have generated estimates of additional health expenditures running into the hundreds of billions of dollars (Coller, Harrison and McInnes (2002)). Evidently, a better understanding of the determinants of smoking behaviour continues to have great significance for health policy.

Smoking involves an intertemporal trade off that should be apparent: any short-term benefits from smoking are coupled with the potential for large long-term costs. In addition, the decision to smoke involves risks that should be apparent, such as the possibility of negative health consequences, and is made under conditions of uncertainty, without the person knowing his or her susceptibility to these risks.

Table 1 provides a detailed summary of experimental studies investigating the relationship between smoking and time preferences. Online searches of PubMed and Econlit, employing the search criteria "smoking" and "discounting" and their variants (e.g., "smoke", "discount", and "time preference"), were used to locate these studies. An initial list of over 50 studies was trimmed according to the following criteria: the study had to include a clear smoker, non-smoker comparison[3]; study participants had to make choices between amounts of real money, rather than cigarettes or quality-adjusted life years, available at different points in time[4]; and the instrument used to

[3] A number of studies (e.g., Field et al. (2006), Dallery and Raiff (2007), Epstein et al. (2003)) focus purely on discounting among smokers and were excluded due to the lack of non-smokers in the sample.
[4] Odum, Madden and Bickel (2002) and van der Pol and Ruggeri (2008) focus on the discounting of health outcomes and Field et al. (2006) and Odum and Baumann (2007) focus on the discounting of hypothetical cigarette rewards.

assess discounting had to include at least 20 questions.[5] The 31 studies satisfying our inclusion criteria are listed in Table 1; a detailed discussion of this table is provided in Appendix A.

[Table 1 here]

The last column of Table 1 reports whether a significant statistical relationship was found between smoking and discounting behaviour. A "positive" relationship between smoking and discounting means that smokers discount more heavily than non-smokers, consistent with expectations before the reported observations. Some of the entries in Table 1 report findings from several studies or from different treatments in the same study. For example, Baker, Johnson and Bickel (2003) report results from real and hypothetical experimental treatments whereas Chabris et al. (2008) report findings from multiple studies. In some cases (e.g., Baker, Johnson and Bickel (2003)) results were the same across studies and treatments, while in others (e.g., Chabris et al. (2008), Heyman and Gibb (2006)) they differed. The last column of Table 1 therefore summarises the set of 37 reported findings from the 31 studies.

Of the 37 reported findings in Table 1, 29 were positive and significant while the remaining 8 were null results.[6] Thus, the bulk of findings in this literature point to a positive relationship between smoking and greater discounting behaviour, irrespective of whether real or hypothetical rewards, long or short temporal horizons, choice or titration elicitation mechanisms, small or large samples, and simple or complex statistical procedures were used.

[5] Some panel studies, such as the Health and Retirement Study (HRS), include a module to assess discounting behaviour but the limited number of questions (e.g., three questions in the HRS, see Bradford (2010)) makes precise estimation and inference difficult. Hence these studies were excluded.

[6] Some studies classified smokers using more than one category (e.g., heavy and light smokers in Stillwell and Tunney (2012)), others classified non-smokers using more than one category (e.g., never-smokers and ex-smokers in Bickel, Odum and Madden (1999)), and still others separated male and female smokers and non-smokers (e.g., Jones et al. (2009) and HLR). In a few of these cases, comparisons between some of the groups were significant while others were not, which makes coding the study problematic. Studies were therefore coded as having found a significant result if at least one smoker, non-smoker comparison was statistically significant. This procedure is preferable to coding a study as having found no statistically significant results just because one comparison (between, say, light smokers and non-smokers) was not significant even though another comparison (between, say, heavy smokers and non-smokers) was significant.

From a statistical perspective, the most striking feature of Table 1 is the near-universal two-step approach to data analysis. This approach entails using nonlinear least squares (NLLS), or some similar technique, to estimate discounting parameters at the level of the individual, and then using the, typically log-transformed, point estimates as data in subsequent statistical models. Harrison, Lau and Rutström (2010) (HLR) and Hofmeyr et al. (2017) are the only studies in Table 1 which do not use this method. The problem with the two-step approach, aside from typically relying on tiny samples to estimate discounting parameters at the level of the individual, is that estimated discounting parameters are estimates, not data. Such estimates comprise both a point estimate (of the mean) and a standard error, and to use only the point estimate is to throw away information on the sampling variability of that estimate.

Moreover, using an estimated discounting parameter as data violates one of the statistical assumptions of the second-stage models: that the covariates are measured without error. Thus, statistical inferences drawn from this approach are simply invalid. HLR and Hofmeyr et al. (2017) estimate discounting parameters as a linear function of observable characteristics (e.g., age, gender, and smoking status) so that the uncertainty of the discounting parameter estimates propagates into the inferences which are drawn from the data.[7] This valid statistical approach will be used here.

Table 2 provides a detailed summary of studies investigating the relationship between smoking and risk preferences. Unlike the literature on time preferences and smoking behaviour, there is a dearth of studies analysing the risk preferences of smokers and non-smokers. Online searches of PubMed and Econlit, employing the search criteria "smoking" and "risk preferences" and their variants (e.g., "smoke," "risk", and "probability discounting"), were used to locate these papers. An initial list of studies was trimmed according to the following rules: the study had to include a clear

[7] To explain the importance of this approach, suppose that the point estimates of a discounting parameter are higher, on average, for smokers than non-smokers. But assume that the estimates of this discounting parameter have high noise (viz., standard errors). Comparing only the signals (viz., point estimates) may lead one to erroneously conclude that smokers discount at a significantly higher rate than non-smokers when an analysis that incorporates both the signals and the noise would find no significant difference between the groups. The method we adopt incorporates both the signals and the noise so that valid inferences can be drawn.

smoker, non-smoker comparison[8]; and study participants had to have made choices between lotteries[9] involving amounts of money, rather than cigarettes or quality-adjusted life years.[10] The 11 studies satisfying our inclusion criteria are listed in Table 2; a detailed discussion of this table is deferred to Appendix B.

Table 2 shows that a majority of the studies (8 out of 11) adopted the probability discounting (PD) approach to risk preferences, which defines risk aversion solely in terms of the shape of the probability weighting function (PWF).[11] The PD model is just Yaari's (1987) dual theory of choice under risk limited to a circumscribed class of lotteries and with a specific PWF: $\pi(p) = p \,/\, [p + \gamma(1 - p)]$. If $\gamma > 1$ this specification represents probability pessimism and risk aversion. As subjective probability distortions drive risk preferences in the PD framework, it is surprising that 6 out of these 8 studies only used 5 probabilities in the elicitation task; the remaining two studies (Mitchell (1999) and Yi, Chase and Bickel (2007)) only used 6 and 7 probabilities, respectively.

[Table 2 here]

The final column of Table 2 shows whether the studies found a significant statistical relationship between risk preferences and smoking behaviour: the results are equivocal and, other than HLR, the statistical analyses are not valid. A positive relationship between smoking and risk preferences means that smokers are more risk averse than non-smokers, whereas a negative relationship means that smokers are less risk averse than non-smokers. Null results were reported in 3 studies, positive results were reported in 5 studies, and negative results were reported in 3 studies.[12] These

[8] Lawyer et al. (2011) investigate whether the risk preferences of smokers and non-smokers differ when they make choices over hypothetical or real rewards. However, they do not compare the risk preferences of smokers and non-smokers.
[9] A number of studies (e.g., Bradford (2010), Jusot and Khlat (2013)) use survey questions which try to elicit general attitudes toward risk and were excluded for this reason.
[10] van der Pol and Ruggeri (2008) investigate risk preferences over hypothetical health outcomes.
[11] Of these studies, 3 also employed the area under the curve (AUC) method of Myerson, Green and Warusawitharana (2001). When using the AUC method, one calculates the area under a subject's derived certainty equivalents and normalizes this to lie in the closed unit interval. Larger AUCs imply less risk aversion and, thus, the AUCs of smokers and non-smokers can be compared to determine whether the groups differ in their risk preferences.
[12] Some studies classified smokers using more than one category (e.g., heavy smokers and light smokers in Poltavski and Weatherly (2013), and smokers and "triers" in Reynolds et al. (2003)), and HLR separated male and female smokers and non-smokers. We again adopt the classification scheme

conflicting results cut across different elicitation mechanisms, real and hypothetical rewards, different frameworks for choice under risk, and different methods of analysis. Thus Table 2 shows that the relationship between risk preferences and smoking behaviour, or lack thereof, differs markedly across studies.

Table 2 also shows that every study except HLR again adopted a two-step approach to statistical analysis: NLLS is used to estimate risk preference parameters at the level of the individual and then these point estimates are used as data in subsequent statistical models. For the reasons outlined above, this approach is statistically invalid.

We add to the extant literature by simultaneously investigating the relationship between risk preferences, time preferences and smoking behaviour using an incentive-compatible experimental design, a relatively large sample of South African university students, and a statistical framework which allows one to draw robust inferences about smokers and non-smokers.

IV. EXPERIMENTAL DESIGN AND SUMMARY STATISTICS

We recruited 175 subjects from undergraduate classes at the University of Cape Town (UCT). Given the focus on smoking behaviour, sign-up sheets included a simple screening question: "Do you smoke cigarettes (Yes / No)." A pool of over 900 people applied to take part in the study and individuals from the smoking and non-smoking groups were randomly selected for inclusion in the project. Those who were selected were added to a website which allowed them to sign up for an experimental session that did not conflict with their academic timetable.

The experiment took place in a computer lab at UCT which had been set up to run the risk and time preference software developed by us. Subjects were separated by partitions and were not allowed to talk to each other during the session. The experiment was conducted in August 2012 across 10 sessions. The median group size was 17 participants and one of us assumed the role of experimenter for every session;

that codes a study as having found a statistically significant result if at least one smoker, non-smoker comparison was significant, even if all comparisons were not.

two research assistants (RAs) were also employed to help administer subject payments and answer questions.

Upon arrival at the lab, subjects were randomly allocated to computer terminals and given an overview of the tasks that they would complete. Subjects then signed informed consent before being taken through a detailed presentation of the risk or time preference task.[13] The order of these tasks was counter-balanced across sessions so subjects either performed the risk or time preference task first. Participants were given the opportunity to ask questions at any stage of the presentations or during the tasks. After questions had been addressed, subjects completed the first task.

Once all participants had completed the first task, the experimenter went through a detailed presentation of the other task. Subjects then completed this task before filling out a questionnaire which collected standard demographic characteristics and information on smoking behaviour. The experimenter or RAs then determined their earnings for the tasks. All subjects received a show-up fee of R20. Earnings for the risk preference task were paid out immediately in cash and earnings for the time preference task were paid out on the date corresponding to the subject's choice on the randomly selected discounting question. Delayed payments were effected via electronic transfer and subjects received a payment notification on their cell phones as soon as the transfer took place. Such transfers are a common means of payment in South Africa and were used to reduce the transaction costs which subjects would have had to incur by coming to collect their delayed payments from us. Experimental sessions lasted approximately an hour and subjects earned R370 (roughly $66 at purchasing power parity (PPP) at the time) on average.

A. Risk Preference Task

The risk preference interface was based on Hey and Orme (1994). It presented subjects with a choice between two lotteries on a screen, displayed as pie charts with accompanying text that listed the probabilities of the prizes. Figure 1 shows a

[13] The introductory presentation, the risk preference task presentation, and the time preference task presentation are included in Appendix C. The presentations were designed to make the tasks transparent and easy to understand. The payment system was also discussed in detail so that subjects understood how their final earnings were determined. This attention to detail, coupled with salient rewards, promotes incentive compatibility and the truthful revelation of preferences.

screenshot of the risk preference task. The display seen by subjects used colours, allowing for greater discrimination than might be apparent from a monochrome presentation (e.g., for the Right lottery).

[Figure 1 here]

The task used prize magnitudes between R0 and R280 (roughly $0 to $50 at PPP at the time) and probabilities which varied in increments of 0.05 between 0 and 1. Thus, other than HLR, this study used larger lottery prizes than any of the studies in Table 2 which have incentive-compatible experimental designs. In addition, this study had more variation in the probability domain than every other study in Table 2. This variation provides for enhanced sensitivity to any probability weighting that might be present.

The lottery pairs in the task were based on the set developed by Loomes and Sugden (1998) (LS) to test different stochastic specifications of choice under risk. LS designed the lottery pairs to accommodate a wide range of risk preferences, to provide good coverage of the probability space, and to generate common-ratio tests of expected utility (EU) theory. However, all the lotteries over which each subject made choices had the same context (i.e., the same set of prizes).[14] By contrast, we used four prize contexts in the experiment: (R0, R140, R280), (R40, R80, R240), (R20, R100, R220), and (R60, R120, R180). Incorporating a number of different prizes and probabilities is helpful for the separate identification of the utility function and the PWF in models which admit both sources of risk preferences (e.g., rank-dependent utility theory).

[Figure 2 here]

Figure 2 shows the set of Marschak-Machina (MM) triangles representing the lotteries, and lottery pairs, which were used in the risk preference task. The top of each diagram lists the context of the lotteries (e.g., (R0, R140, R280)) and the

[14] LS used two experimental treatments: one where subjects made choices over lotteries defined on the context ($0, $10, $20) and one where subjects made choices over lotteries defined on the context ($0, $10, $30). The probability distributions over these contexts were identical across the two groups except for 8 out of the 45 lotteries in the task.

gradient of the lines connecting lottery pairs. Each point in the MM triangle represents a lottery and the line connecting two, or more, points represents a lottery pair, or set of lottery pairs, on offer in the choice task. Figure 2 shows that the risk preference task provided thorough coverage of the MM triangle, in the sense of including a combination of interior and boundary choices, and that it captures the full range of risk preferences, under the null hypothesis of EU theory: risk-loving (gradients less than 1), risk neutral (gradients equal to 1), and risk averse (gradients greater than 1). Subjects made 40 choices in the risk preference task and one choice was selected at random at the end of the experimental session for payment.[15]

B. Time Preference Task

The time preference task presented subjects with choices between smaller, sooner (SS) and larger, later (LL) monetary rewards. Figure 3 shows a screenshot of the time preference task. On each screen subjects had to make 4 choices before proceeding to the next screen. The principal (i.e., SS reward) and time horizon were fixed on each screen but varied across screens. A calendar was displayed on every screen to show the subjects when they would receive the amounts of money they chose.

[Figure 3 here]

Following Coller and Williams (1999), three front end delays (FEDs) to the SS rewards were used: zero days, 7 days, and 14 days. This design allows one to hold subjective transaction costs constant for the SS and LL rewards at positive FEDs. It also facilitates estimation of the parameters of a quasi-hyperbolic or β-δ discounting function because the zero day FED allows one to pin down the estimate of β, which captures a "passion for the present" or "present-bias" in decision making, whereas the

[15] We decided to use the Random Lottery Incentive Mechanism (RLIM), where one of the 40 choices was chosen at random to be played out for payment. We did so to ensure that we collected enough choices over a wide enough array of lotteries to be able to identify EU and rank-dependent utility models. If we had opted for giving one choice to each subject, to avoid using RLIM, this would have been infeasible. Harrison and Swarthout (2014) find that using RLIM does make a difference behaviourally when estimating non-EU models, but not, as one would expect, when estimating EU models. A logical response to this problem is simply to assume two independence axioms: one axiom that applies to the evaluation of a given prospect, and that is assumed to be violated by non-EU models, and another axiom that applies to the evaluation of the experimental payment protocol. One can then allow for failure of the former axiom, when estimating non-EU models, but assume the validity of the latter axiom. Cox, Sadiraj and Schmidt (2015) also consider the implications of assuming RLIM, and discuss in detail the strengths and weaknesses of alternative payment protocols.

positive FEDs allow one to estimate the long-term discounting parameter δ. Subjects in an experimental session were only exposed to one of these FED treatments.

Two principals (R150 and R250: roughly $27 and $45 at PPP at the time), 14 time horizons between the SS and LL rewards (7 to 98 days, in 7-day increments), and nominal annual interest rates between 5% and 250% were used in the time preference task. These parameters define a battery of 224 possible choice pairs. Each subject made 60 choices in the task which were drawn randomly, without replacement, from this battery. At the end of the experimental session, one of these choices was randomly selected for payment.

C. Summary Statistics

Table 3 presents summary statistics for the sample of 175 students. The average age in the sample is approximately 20 years old, 42% of the sample is White[16], two-thirds are enrolled in the Commerce faculty at UCT, and approximately one-third receives financial aid. Smokers were defined as those people who answered "yes" to the question: "Do you currently smoke cigarettes?"[17] Current smokers make up 62% of the sample[18] and this is the largest number of smokers (i.e., 108 smokers) ever recruited for a study exploring risk preferences and smoking behaviour. Smokers were deliberately oversampled to investigate whether intensity of smoking is related

[16] Designation of population groups or 'races' follows the traditional categorisation in South Africa that is still employed in affirmative action and related policies, notwithstanding recognition that it involves cultural and historical discriminations that are without biological significance. Approximately 24% of the sample is Black. 14% is Coloured, a culturally salient population group in South Africa composed of individuals of mainly Indonesian descent who speak Afrikaans as a first language. 17% is Indian. The remaining 3% preferred not to classify their race.

[17] There is a vast literature comparing self-reports of smoking with objective measures, e.g., cotinine measures, that are known to be correlated with exposure to smoke (see Gorber et al. (2009) for a survey). Two recent examples come from the Canadian Health Measures Survey (CHMS) and the National Health and Nutrition Examination Survey (NHANES). In the 2007-2009 wave of the CHMS, "ever smokers" were asked detailed questions about their current and recent smoking behavior, and urine cotinine measurements were taken between one day and 6 weeks after the initial survey response. Using these data, Wong et al. (2012) show high levels of consistency between self-reports and objective measures and conclude that, "Representative data for the Canadian population showed no significant difference between national estimates of smoking prevalence based on self-report versus urinary cotinine concentration." Choi and Cawley (2017) reach a similar conclusion using NHANES data from 1999-2012 and find, in addition, that accuracy of self-reported smoking tends to increase with level of education. To the extent that this finding is robust, our university sample of smokers are likely to have given accurate self-reports of smoking.

[18] The remaining 38% of the sample comprises both former-smokers and never-smokers who will be referred to collectively as non-smokers.

to risk and time preferences. The mean number of cigarettes smoked per day is 8.67 with a standard deviation of 5.81 and a range of 1 to 25.[19,20]

[Table 3 here]

Smokers also completed the Fagerström Test for Nicotine Dependence (FTND) due to Heatherton et al. (1991). The FTND is a measure of smoking severity that scores people on a scale of 0 to 10, where higher numbers indicate greater severity. The average FTND score among smokers is 2.22 with a standard deviation of 2.08. Thus, on average, the smokers in this sample are relatively light smokers. In addition, given the young age of the sample, the smokers' lifetime exposure to cigarettes is relatively low. In the literature on risk preferences, time preferences and smoking behaviour, researchers often try to maximise the difference between smokers and non-smokers by selecting heavy smokers to take part in the study. We recruited smokers across the entire spectrum of severity to determine whether being a smoker, irrespective of intensity, is associated with risk and time preferences. This also allows us to explore the relationship between risk preferences, time preferences and smoking intensity.

Table 3 shows that randomisation across experimental treatments ensured that approximately 50% of the sample completed the risk preference task prior to the time preference task. FED treatments were split evenly across the sample and 50% of choices in the time preference task involved the high principal of R250.

[19] Estimates from the South African National Health and Nutrition Examination Survey of the mean number of cigarettes smoked per day for people aged 15-24 is 5.9 (Shisana et al. (2013, p. 111)). For the population as a whole, the mean number of cigarettes smoked per day is 8.5. Thus, our sample, at least in terms of the mean number of cigarettes smoked per day, is very similar to the general population.

[20] According to The Tobacco Atlas (see www.tobaccoatlas.org and Eriksen et al. (2015)), 22.2% of men and 9% of women smoke tobacco daily in South Africa. The prevalence rate for men is lower than in other middle-income countries but the prevalence rate for women is higher than in other middle-income countries. Prevalence rates for selected high-income countries are: US – men: 17.2%, women: 14.2%; UK – men: 23.2%, women: 20.3%; Australia – men: 15.1%, women: 11.6% ; Germany – men: 28%, women: 22.2%.

V. STATISTICAL SPECIFICATION

The statistical method we employ is direct estimation by maximum likelihood of structural models of latent choice processes. The latent choice processes in question are captured by models of risk and time preferences. These models provide the structure necessary to estimate risk and time preferences using the observed choice data. One of the benefits of the maximum likelihood approach is that it uses all of the available information to estimate discounting and risk preference parameters and the precision of these estimates. This estimation strategy closely follows Andersen et al. (2008) and HLR so we provide a brief explanation of the method below, focussing on the canonical cases of EU theory and exponential (E) discounting. Further details are provided in Appendix D. We also discuss the extension to other risk and time preference models.

Assume that utility of income is defined by a power utility function which displays constant relative risk aversion (CRRA):

$$U(y) = y^r, \tag{1}$$

where y is a lottery prize in the risk preference task and r is a parameter to be estimated.

To estimate the parameter r we formed a latent index, based on latent preferences, that captured the difference in the expected utility of the Right and Left lotteries presented to subjects. The value of this index, for each observation, was determined by the lottery prizes, their associated probabilities, and an initial estimate of r. This latent index was linked to the subjects' binary choices (i.e., the Left or Right lottery) using the cumulative normal distribution function. This "probit" link function determined the likelihood of selecting the Left lottery, and hence the likelihood of selecting the Right lottery, for each observation in the dataset given the value of the latent index. Maximum likelihood estimation was then used to determine the value of r that maximised the likelihood of observing all of the data from the experiment.

It is a straightforward extension to make the parameter r a linear function of individual characteristics in order to draw robust inferences about potential differences in the risk preferences of participants. In addition, every estimate of r

includes a standard error which reflects our uncertainty as to the "true" value of *r*. This stands in sharp contrast to the bulk of studies in Table 2 which use risk preference point estimates as data in subsequent statistical models. We also extended the model by adopting the "contextual utility" (CU) behavioural error specification of Wilcox (2011) to allow mistakes on the part of subjects from the perspective of the deterministic EU model and to draw robust inferences about the primitive "stochastically more risk averse than" relation.[21]

It is a simple matter to incorporate other theories of choice under risk in this statistical framework. Quiggin (1982) developed the rank-dependent utility (RDU) model, which assumes that a decision maker transforms objective probabilities into subjective decision weights which are then used to evaluate lotteries. In this context, we estimate the parameters of a utility function and PWF which maximise the likelihood of observing the data from the experiment on the basis of a latent index which captures the difference in the rank-dependent utility of the lotteries.

We estimate EU and RDU models to compare the risk preferences of smokers and non-smokers. In addition, we estimate the parameters of a variety of PWFs to ensure that the results are robust across different specifications.

Shifting to time preferences, under the E model, δ is the discounting parameter which equalises the *utility* of income received at time t (i.e., the utility of the SS reward) with the *utility* of income received at time $t + \tau$ (i.e., the utility of the LL reward):

$$[1 / (1 + \delta)^t]U(y_t) = [1 / (1 + \delta)^{t+\tau}]U(y_{t+\tau}), \tag{2}$$

for some utility function $U(\cdot)$.

Under the assumptions that EU characterises choices over risky prospects and that subjects employ the power utility function, we can add more structure to this indifference condition. Specifically, (2) becomes:

$$[1 / (1 + \delta)^t](y_t)^r = [1 / (1 + \delta)^{t+\tau}](y_{t+\tau})^r, \tag{3}$$

where the general form of the utility function $U(\cdot)$ in (2) has been replaced with the specific power utility function $U(y) = y^r$ in (3).

[21] The "stochastically more risk averse than" relation is the stochastic choice counterpart to the "more risk averse than" relation (see Pratt (1964)) which is defined for the deterministic EU model.

To estimate the parameters of our time preference model, conditional on EU theory, power utility, and the E model, we form a latent index that captures the difference in the present value of the utility of the SS and LL rewards, and we incorporate the behavioural error term originally due to Fechner (1966/1860).

This "joint estimation" approach, developed by Andersen et al. (2008), uses subjects' choices in the risk preference task to pin down the parameters of the utility function, and subjects' choices in the time preference task to pin down the parameters of the E discounting model, conditional on the shape of the utility function. This approach ensures that we estimate time preferences defined over utility flows, and not flows of money.

It is straightforward to incorporate other discounting models in this statistical framework. In the case of Weibull discounting, for instance, (3) becomes:

$$[\exp(-\delta t^{(1/\beta)})](y_t)^r = [\exp(-\delta(t+\tau)^{(1/\beta)})](y_{t+\tau})^r \qquad (4)$$

We then form the latent index that captures the difference in the present value of the utility of the SS and LL rewards and proceed as before.

VI. RESULTS

We present the results from a set of risk and time preference models so as to explore the relationship between risk preferences, time preferences and smoking behaviour. We begin with the risk preference results because they provide a natural segue to the time preference results which are conditional on the utility function curvature identified by the risk preference task.

A. Risk Preferences

We estimate an EU model employing a power utility function and the CU behavioural error specification; see Appendix E for more details. We find a relatively high level of risk aversion in the sample; a statistically significant estimate of the behavioural error parameter, implying that subjects make behavioural errors in the risk preference task; and no substantive differences in the risk preferences of smokers and non-smokers. We also estimate a model which allows risk preferences to vary as a quadratic function of smoking intensity as measured by the number of cigarettes smoked per day: risk preferences are not significantly related to smoking intensity. These results are robust to the assumption that Saha's (1993) expo-power utility function – which admits increasing relative risk aversion, decreasing relative risk aversion, and CRRA – characterises choice under risk.

The EU results suggest that there are no significant differences in the risk preferences of smokers and non-smokers. However, this analysis, by assumption, ignores the role of probability weighting and it may be the case that smokers perceive probabilities differently to non-smokers. For example, smokers may underweight moderate to high probabilities more so than non-smokers, and may, therefore, underestimate the likelihood of the negative consequences associated with smoking. To explore this possibility, we estimate RDU models.

One of the key components of a RDU model is the specification of the PWF. We estimate the power PWF, the PWF popularised by Tversky and Kahneman (1992) (TK), and the Prelec (1998) two-parameter PWF which exhibits considerable flexibility; see Appendix D for more details. The functional form for the Prelec (1998) PWF is:

$$\pi(p) = \exp[-\eta(-\ln p)^{\varphi}], \qquad (5)$$

which is defined for $1 > p > 0$, $\eta > 0$, and $\varphi > 0$. This function allows independent specification of location and curvature in probability weighting. It also nests the power PWF when $\varphi = 1$, and nests a one-parameter function when $\eta = 1$, which is similar to the TK function and admits linear, inverse S-shaped, and S-shaped forms.

We find statistically significant evidence of inverse S-shaped probability weighting. To investigate the possibility that smokers perceive probabilities differently to non-smokers we estimate a RDU model with a power utility function, the CU behavioural error specification, and the Prelec (1998) PWF, and allow the parameters to vary as a function of observable characteristics and task parameters. Results are presented in Table 4.[22] Smokers do not differ significantly from non-smokers in the shape of their utility functions (i.e., in the estimate of r) nor in the way they perceive probabilities

[22] Appendix E also presents results from a RDU model employing the TK PWF; the results are qualitatively identical to those in Table 4.

(i.e., in the estimates of φ and η). In addition, tests of the joint hypothesis that the coefficients for smokers across *r*, φ, and η are equal to zero, cannot be rejected (p = 0.823).[23,24]

[Table 4 here]

Thus, at least in this sample, there are no significant differences in risk preferences according to smoking behaviour. This result is robust to different theories of choice under risk, different PWFs, and a utility function that admits varying relative risk aversion.

B. Time Preferences

We estimate four time preference models: the E model, the quasi-hyperbolic (QH) model, Mazur's (1984) hyperbolic (H) model, and the Weibull (WB) model; see Appendix F for more details and Andersen et al. (2014) for a review of all of the major discounting models. We employ a Fechner error term and jointly estimate the parameters of these models with the curvature of the utility function, assuming RDU[25] and the Prelec (1998) PWF characterise choice under risk, to focus on the discounting of utility flows, not flows of money. In the context of addiction, the crucial difference between these time preference specifications is that, under the assumption of an additively-separable intertemporal utility function, the E model implies time-consistent preferences whereas the other models can yield time-inconsistent preferences.[26]

[23] We also estimate a RDU model with the expo-power utility function, the Prelec (1998) PWF, and the full set of covariates from Table 4. The smoker variable is not significantly different from zero for any of the parameters in the model. In addition, a test of the joint hypothesis that the coefficients for smokers across *r*, α, φ, and η are equal to zero, cannot be rejected (p = 0.967).

[24] We also investigate the relationship between smoking intensity and risk preferences by estimating the model in Table 4 and allowing the parameters of interest to vary as a quadratic function of number of cigarettes smoked per day. None of the linear or quadratic terms are statistically significant in any of the equations and a joint test of the linear *and* quadratic terms across all equations is not statistically significant either (p = 0.576).

[25] Given the presence of probability weighting in this dataset, we employ RDU theory to apportion risk preferences into their concave utility and probability weighting components so as to draw accurate inferences about discounting behaviour. If one ignores probability weighting when it is present, this would lead to biased estimates of utility function curvature and, hence, biased estimates of discounting parameters. In Appendix G we test the robustness of our results by estimating these models assuming EU theory characterises choice under risk; the results are *qualitatively* identical to those reported in the main text.

[26] Time consistency, or the lack thereof, is central to economic models of addiction. Time-inconsistent agents may fail to carry out plans they make for the future which provides a possible explanation for

The estimate of the E discount rate $\delta = 0.493$ implies an annual discount rate of approximately 49%, which is a marked decline in comparison to the estimate of $\delta = 3.234$ under the assumption of linear utility (see Appendix F). Similar declines are evident across all of the discounting specifications which highlights the point, now familiar from Andersen et al. (2008), that incorporating concavity of the utility function leads to substantial declines in inferred discount rates.

In the QH model, the estimate of $\beta = 0.988$, which captures a "present-bias" or a "passion for the present" in discounting behaviour, is statistically significantly less than 1 ($p = 0.002$), which provides evidence of quasi-hyperbolic discounting and declining discount rates. The same is true in the WB results: the estimate of $\beta = 1.611$, which "expands" or "contracts" time, is statistically significantly greater than 1 ($p < 0.001$) which leads us to infer that people perceive time as "slowing down," generating declining discount rates. Thus, both the QH and WB results suggest that discount rates decline over time, which, when coupled with an additively-separable intertemporal utility function, raises the spectre of time-inconsistent choices. However, the two discounting functions provide competing explanations for this result: a present-bias in the case of the QH model and subjective time perception in the case of the WB model.

As a descriptive prelude to the formal statistical results, Figure 4 shows a kernel-weighted local polynomial regression, with a 95% confidence interval, of the fraction of LL choices by smokers and non-smokers for the nominal annual interest rates on offer in the time preference task. At each interest rate, the point estimate of the fraction of LL choices by smokers is less than the point estimate of the fraction of LL choices by non-smokers, and the 95% confidence intervals do not overlap. This suggests that smokers discount more heavily than non-smokers, but clearly this result must be subjected to closer statistical scrutiny before any definitive conclusions are reached.

---

the behavioural puzzles listed earlier: addicts expend resources to acquire their targets of addiction but then incur real costs to try to reduce or limit their consumption of these goods; and the fact that the typical course of addiction is characterised by repeated unsuccessful attempts to quit prior to final abstention.

[Figure 4 here]

Consequently, we estimate the four time preference models, assuming RDU and the Prelec (1998) PWF, where risk and discounting parameters are allowed to vary by smoking status, other observable characteristics, and task parameters; see Appendix F for the results. Across all specifications, the effect of smoking on the estimate of δ is positive and statistically significant at the 1% level, implying that smokers tend to discount the future more heavily than non-smokers. The magnitude of this difference in discounting behaviour is economically significant. In the E model, for example, smokers have an annual discount rate which is 26 percentage points higher than non-smokers. Thus, the positive relationship between smoking and discounting identified in Table 1 has been replicated using a full set of covariates and a joint estimation approach to time preferences which controls for utility function curvature and probability weighting.[27]

The estimates of β in the QH and WB models, by contrast, do not vary according to smoking status. Thus, smokers are no more present-biased than non-smokers in the QH model nor are they more likely to perceive time as slowing down in the WB model. It is only the long-term discount rate δ which differs between smokers and non-smokers in these models.

It is inferentially risky to try to boil down smoking to a binary covariate (e.g., smoker, non-smoker) because one runs the risk of mischaracterising the full effects of smoking behaviour if there are differences between non-smokers, light smokers, moderate smokers, and heavy smokers. Thus, to extend our analysis, we investigate whether smoking intensity and discounting behaviour are related by estimating the four time preference models and allowing the parameters of interest to vary as a quadratic function of the number of cigarettes smoked per day, other observable characteristics, and task parameters.

[27] Appendix G presents results from the four time preference models where EU theory is assumed to characterise choice under risk: the results are virtually identical to the models which assume RDU and the Prelec (1998) PWF.

In all models, both the linear and quadratic terms are statistically significant in the estimate of δ: the linear term is positive and significant whereas the quadratic term is negative and significant. Thus, there is a concave relationship between discounting behaviour and number of cigarettes smoked per day: every additional cigarette is associated with an increase in discounting, but at a decreasing rate until a maximum is reached, after which every additional cigarette is associated with a decrease in discounting.[28]

[Table 5 here]

Table 5 maps out the response surface for estimates of δ in the four time preference models evaluated at different values of number of cigarettes smoked per day. At low values of number of cigarettes, the conditional marginal effect of additional cigarettes is positive. By 15 cigarettes, though, the conditional marginal effect of additional cigarettes is negative. Table 5 highlights the nonlinear effect of smoking intensity on discounting behaviour. To our knowledge, this is the first study of time preferences and smoking behaviour which has identified this effect.

C. Mixture Models of Discounting Behaviour

The analyses conducted thus far have been based on the implicit assumption that the observations are produced by only one discounting data generating process (DGP): either E, H, QH, or WB. However, the data may be a result of more than one DGP. For example, the E model may explain some discounting choices better than the H model whereas the H model may explain other choices better than the E model. The assumption that only one DGP characterises all of the data precludes such a possibility.

Finite mixture models[29] allow two or more DGPs to account for the data and also provide a measure of the proportion of the data which is best explained by each

[28] In the QH model, smoking intensity is not significantly related to the extent of present-bias. In the WB model, though, the number of cigarettes' linear term is negative and significant in the estimate of β, albeit at the 10% level. Thus, the more cigarettes smoked per day, the less likely people are to perceive time as slowing down.

[29] For detailed discussions of mixture models see McLachlan and Peel (2000), Harrison and Rutström (2009), and Conte, Hey and Moffatt (2011). Mixture models have been applied to discounting behaviour by Andersen et al. (2008), Coller, Harrison and Rutström (2012) and Andersen et al. (2014).

process. In the current context, one can estimate a mixture model of, say, the E and H discounting functions and then ask the data to determine each function's level of support. To do so one specifies a "grand likelihood" function which is just a probability-weighted average of the likelihoods of the two models; see Appendix H for more details.

We estimate a mixture model of the E and H discounting functions and both functions find statistically significant support in the data.[30] In addition, the mixture model shows that discounting parameter estimates are distorted when the E or H models have to account for all of the data. We also use the mixture model to explore the factors that may affect the likelihood of discounting according to the E and H functions.

Given that the typical pattern of addiction is characterised by choice behaviour that implies time-inconsistent preferences, it is of particular importance to determine whether smoking behaviour is associated with a greater likelihood of discounting according to the time-inconsistent H model as opposed to the time-consistent E model. Given our interest in smoking intensity, rather than a binary classification of smoking status, we estimate a mixture model of the E and H discounting functions and allow the risk and time preference parameters to vary as a quadratic function of number of cigarettes smoked per day. In addition, we include a full set of covariates in the mixture probability equation and a number of cigarettes smoked per day linear term[31] to identify the factors that may affect the likelihood of discounting according to the E and H models.

[Table 6 here]

[30] Appendix H contains the results from all of the two-process mixture models that can be estimated from the four discounting specifications. We only discuss the results from the E and H mixture model in this section because these are the most commonly used discounting functions in the addiction literature and they are representative of the results from the other mixture models.

[31] We also estimate the mixture model with a full set of covariates in the mixture probability equation and allow this equation to vary as a quadratic function of number of cigarettes smoked per day. The quadratic term is not statistically significant, implying that we do not need to incorporate higher order polynomials of this variable in the equation, and can employ the linear term by itself. As would be expected from the results in Table 6, when we incorporate both the linear and quadratic terms of number of cigarettes smoked per day, a joint test of the coefficients on these terms is statistically significant at the 5% level.

Table 6 presents the results. Of particular interest is that the number of cigarettes smoked per day is negatively and statistically significantly ($p < 0.01$) related to the likelihood of discounting according to the E model. The magnitude of this effect is large: every additional cigarette smoked per day is associated with a 2 percentage point *decrease* in the likelihood of discounting according to the E model and, hence, a 2 percentage point *increase* in the likelihood of discounting according to the H model. Thus, as smoking intensity increases this is associated with a greater likelihood of discounting hyperbolically as opposed to discounting exponentially. This result has important implications for our understanding of addiction. In addition to the nonlinear effect of smoking intensity on inferred discount rates identified in the previous section, smoking intensity is also linked to the likelihood of making time-inconsistent choices, which is the hallmark of addictive consumption patterns. To our knowledge, this is the first study to have identified this effect.

VII. DISCUSSION AND CONCLUSIONS

We analyse the relationship between risk preferences, time preferences and smoking behaviour using an incentive-compatible experimental design and a joint estimation approach to data analysis. We find that both probability weighting and utility function curvature affect attitudes to risk in this sample but we find no statistically significant relationship between risk preferences and smoking behaviour. This result is robust to different theories of choice under risk, different PWFs, and different utility functions which admit varying relative risk aversion.

To analyse the time preferences of our sample we adopt the methodology of HLR which jointly estimates utility function curvature and discounting functions so as to characterise time preferences over utility flows, not flows of money. We find that controlling for the concavity of the utility function leads to a dramatic decline in estimates of δ, replicating the result of Andersen et al. (2008). We also allow RDU to characterise choice under risk so as to apportion risk preferences into their utility curvature and probability weighting components.

We explore the relationship between time preferences and smoking behaviour in three ways. First, we focus on the marginal effect of smoking status on time preferences by estimating the discounting models and making the parameters of interest a linear function of observable characteristics and task parameters. Across all specifications, the estimate of δ for smokers is positive and statistically significant, implying that smokers discount at a higher rate than non-smokers. In Appendix G we also test to see whether these results are robust to the assumption that EU characterises choice under risk: the results are qualitatively identical to those in Section VI.

Second, to investigate whether smoking intensity is related to discounting behaviour, we estimate the four time preference models and allow the parameters of interest to vary as a quadratic function of number of cigarettes smoked per day, other observable characteristics, and task parameters. These analyses reveal a concave relationship between smoking intensity and estimates of the discounting parameter δ. Specifically, every additional cigarette is associated with an increase in discounting, but at a decreasing rate until a maximum is reached, after which every additional cigarette is associated with a decrease in discounting.

Finally, we estimate mixture models of the different discounting specifications and focus on the link between smoking intensity and the likelihood of making time-inconsistent choices. We find that smoking intensity is positively and significantly related to the likelihood of discounting hyperbolically, which suggests that smokers, and, in particular, heavier smokers, are more likely to make time-inconsistent choices.

This research makes a number of contributions to the literature. When analysing risk preferences and smoking behaviour, we allow risk attitudes to be determined both by utility function curvature and probability weighting. Prior studies in the literature either focus on utility function curvature or probability weighting, not both. Consequently, they are always open to the critique that the other source of risk attitudes, the one not explored in the study, differs according to smoking behaviour. Incorporating both utility function curvature and probability weighting in estimates of risk attitudes provides us with immunity to this critique and allows us to make stronger claims about differences in risk preferences according to smoking behaviour.

This is only the second study in the smoking-discounting literature to incorporate utility function curvature in the estimation of time preference models, and it is the first which allows RDU to characterise choice under risk. Although the qualitative discounting estimates do not differ significantly across the EU and RDU specifications, it is nevertheless theoretically appropriate to quantitatively apportion risk preferences into their utility curvature and probability weighting components.

This is the first study to identify a nonlinear effect of smoking intensity on discounting behaviour. Smoking more cigarettes is associated with an increase discounting but only up to a point, after which each additional cigarette is associated with lower discounting. This nonlinear effect may explain why some studies, which only recruited heavy smokers and never-smokers, fail to find a difference in discounting behaviour between these groups.

In addition, this nonlinear effect of smoking intensity may provide an explanation for patterns of cigarette consumption. It has long been assumed that the marked modal clustering around 20 cigarettes per day in mature smokers simply reflects the fact that cigarettes are typically sold in packs of 20. It may be the case, though, that cigarette companies learned to sell cigarettes in packs of 20 because that is where the psychofunctional, and not merely the homeostatic, equilibrium lies for the majority of mature smokers.

This research also reiterates the point that multiple decision processes characterise the discounting of delayed rewards. It is crucial for researchers to be cognisant of this fact when exploring the smoking-discounting relationship. Smoking intensity increases the likelihood of discounting hyperbolically, which may be an important factor in tobacco addiction and explain recalcitrance to treatment. To our knowledge, this is the first study in the literature to identify this effect in a sample of smokers and non-smokers.

This research naturally involves some limitations. Clearly our sample of young South African university students is not representative of a general population, and the smokers among them are not representative of smokers in general. But the significance of our findings, we suggest, does not depend on supporting inferences about general populations. Existing theories of addiction focus on differences between addicts and non-addicts. Since people who smoke as few as five cigarettes every day can be addicted, our observation of effects of smoking intensity on key variables in the economic structure of choice is novel. The "clean" conditions of the laboratory often furnish, as here, the best initial environment for detecting effects not predicted by established theory. The next step in follow-up research is obviously to use larger, more representative samples, along with field studies, to determine whether the effects are robust.

Another potential issue with the sample is the extent of possible selection bias. As discussed earlier, a large number of people applied to take part in the study, so people in the smoking and non-smoking groups were randomly selected to form the study pool. It may be the case that those who were selected were not representative of their group. Ideally we would use information on the population of smokers and non-smokers at UCT to correct for any sample selection issues present in the data.[32] Unfortunately, we do not have any additional information on the population of smokers and non-smokers at UCT.

A question that arises naturally in this line of research is whether risk and time preferences are domain- or context-specific. A noteworthy feature of the limited existing empirical literature on addiction and risk and time preferences is that the latter are invariably measured in the domain of responses to monetary rewards, despite the fact that the most directly relevant arguments of utility functions where addiction is concerned refer to social and health status. While it is possible that most people's risk and time preferences are closely related across domains, this cannot be assumed, especially in a population that is already atypical in being characterised by addiction.

It is practically challenging to address the question of cross-domain preference structure consistency in the laboratory using hypothetical rewards because one loses salience and dominance without money as a reward medium when trying to induce

[32] Harrison, Lau and Rutström (2009) and Harrison and Lau (2014) analyse the effect of sample selection bias on estimated risk preference parameters. They used the Danish Registry to gather information on people who were invited to participate in their experiment but who did not take part and this allowed them to make sample selection corrections for the sample of people who were invited and who did participate in the experiment. Harrison, Lau and Rutström (2009) find that correcting for sample selection bias leads to attenuated risk aversion estimates, implying that their sample was more risk averse than the population from which it was drawn. Similarly, Harrison and Lau (2014) find that sample selection corrections lead to lower estimates of risk aversion.

value. Arguably, the best long-run methodology for handling this difficulty will be to use laboratory work on choices over money as a baseline for extensions into the field where participants' choices affect their real health and social well-being. In that case the first stage research involving monetary rewards is the immediate priority.

We stress that our results refer to correlations between smoking behaviour and preferences. It is apparent that causality can run in both directions, even if we have priors that favour the causal effect of preferences on smoking behaviour as being more prominent. There are several ways to go beyond statements about correlation, which should be considered in future work. One is to mimic a randomised control trial, by matching smokers and non-smokers using some metric such as a propensity score (see Rubin (1998, 2001)), and then evaluating the risk and time preferences of these matched samples. This approach avoids the obvious ethical problem of randomising "smoking" to a sample. One problem with this approach is the need for much larger samples than we have available. A more fundamental problem is that it requires that we reduce "smoking" to a coarse representation of the full characterisation of smoking behaviour (e.g., to a binary variable, an ordered discrete variable, or a single continuous variable). This would blunt the very non-linearity of smoking intensity that is one of our major findings.

These issues notwithstanding, we provide a rigorous framework within which to analyse risk preferences, time preferences and smoking behaviour. Future experimental research should abandon binary classifications of smoking status and seek to replicate the nonlinear effect of smoking intensity on discounting behaviour, and the link between smoking intensity and the likelihood of making time-inconsistent choices. If these results hold in other samples, our understanding of smoking specifically, and addiction generally, will be sharpened.

REFERENCES

ADAMS, J., AND D. NETTLE (2009): “Time Perspective, Personality and Smoking, Body Mass, and Physical Activity: An Empirical Study,” *British Journal of Health Psychology*, 14, 83-105.

AINSLIE, G. (1992): *Picoeconomics: The Interaction of Successive Motivational States within the Individual*. Cambridge: Cambridge University Press.

AINSLIE, G. (2001): *Breakdown of Will*. Cambridge: Cambridge University Press.

ANDERSEN, S., G. W. HARRISON, M. I. LAU, AND E. E. RUTSTRÖM (2008): “Eliciting Risk and Time Preferences,” *Econometrica*, 76, 583-618.

ANDERSEN, S., G. W. HARRISON, M. I. LAU, AND E. E. RUTSTRÖM (2014): “Discounting Behavior: A Reconsideration,” *European Economic Review*, 71, 15-33.

ANDERSON, L. R., AND J. M. MELLOR (2008): “Predicting Health Behaviors with an Experimental Measure of Risk Preference,” *Journal of Health Economics*, 27, 1260-1274.

AUDRAIN-MCGOVERN, J., D. RODRIGUEZ, L. H. EPSTEIN, J. CUEVAS, K. RODGERS, AND E. P. WILEYTO (2009): “Does Delay Discounting Play an Etiological Role in Smoking or Is It a Consequence of Smoking?,” *Drug and Alcohol Dependence*, 103, 99-106.

BAKER, F., M. W. JOHNSON, AND W. K. BICKEL (2003): “Delay Discounting in Current and Never-before Cigarette Smokers: Similarities and Differences across Commodity, Sign, and Magnitude,” *Journal of Abnormal Psychology*, 112, 382-92.

BALEVICH, E. C., N. D. WEIN, AND J. D. FLORY (2013): “Cigarette Smoking and Measures of Impulsivity in a College Sample,” *Substance Abuse*, 34, 256-62.

BECKER, G. S., AND K. M. MURPHY (1988): “A Theory of Rational Addiction,” *Journal of Political Economy*, 96, 675-700.

BÉNABOU, R., AND J. TIROLE (2004): “Willpower and Personal Rules,” *Journal of Political Economy*, 112, 848-886.

BENHABIB, J., AND A. BISIN (2004): “Modeling Internal Commitment Mechanisms and Self-Control: A Neuroeconomics Approach to Consumption-Saving Decisions,” *Games and Economic Behavior*, 52, 460-492.

BERNHEIM, B. D., AND A. RANGEL (2004): “Addiction and Cue-Triggered Decision Processes,” *American Economic Review*, 94, 1558 - 1590.

BICKEL, W. K., D. P. JARMOLOWICZ, E. T. MUELLER, C. T. FRANCK, C. CARRIN, AND K. M. GATCHALIAN (2012): “Altruism in Time: Social Temporal Discounting Differentiates Smokers from Problem Drinkers,” *Psychopharmacology*, 224, 109-20.

BICKEL, W. K., A. L. ODUM, AND G. J. MADDEN (1999): “Impulsivity and Cigarette Smoking: Delay Discounting in Current, Never, and Ex-Smokers,” *Psychopharmacology*, 146, 447-54.

BICKEL, W. K., R. YI, B. P. KOWAL, AND K. M. GATCHALIAN (2008): “Cigarette Smokers Discount Past and Future Rewards Symmetrically and More Than Controls: Is Discounting a Measure of Impulsivity?,” *Drug and Alcohol Dependence*, 96, 256-62.

BRADFORD, W. D. (2010): “The Association between Individual Time Preferences and Health Maintenance Habits,” *Medical Decision Making*, 30, 99-112.

BUSINELLE, M. S., M. A. MCVAY, D. KENDZOR, AND A. COPELAND (2010): "A Comparison of Delay Discounting among Smokers, Substance Abusers, and Non-Dependent Controls," *Drug and Alcohol Dependence*, 112, 247-50.

CHABRIS, C., D. LAIBSON, C. L. MORRIS, J. P. SCHULDT, AND D. TAUBINSKY (2008): "Individual Laboratory-Measured Discount Rates Predict Field Behavior," *Journal of Risk and Uncertainty*, 37, 237-269.

CHOI, A., AND J. CAWLEY (2017): "Health Disparities across Education: The Role of Differential Reporting Error," *Health Economics, forthcoming*, DOI: 10.1002/hec.3609.

CLARK, A. (1997): *Being There*. Cambridge, MA: MIT Press.

COLLER, M., G. W. HARRISON, AND M. M. MCINNES (2002): "Evaluating the Tobacco Settlement Damage Awards: Too Much or Not Enough?," *American Journal of Public Health*, 92, 1-6.

COLLER, M., G. W. HARRISON, AND E. E. RUTSTRÖM (2012): "Latent Process Heterogeneity in Discounting Behavior," *Oxford Economic Papers*, 64, 375-391.

COLLER, M., AND M. B. WILLIAMS (1999): "Eliciting Individual Discount Rates," *Experimental Economics*, 2, 107-127.

CONTE, A., J. D. HEY, AND P. G. MOFFAT (2011): "Mixture Models of Choice under Risk," *Journal of Econometrics*, 162, 79-88.

COX, J. C., V. SADIRAJ, AND U. SCHMIDT (2015): "Paradoxes and Mechanisms for Choice under Risk," *Experimental Economics*, 18, 215-250.

DALLERY, J., AND B. R. RAIFF (2007): "Delay Discounting Predicts Cigarette Smoking in a Laboratory Model of Abstinence Reinforcement," *Psychopharmacology*, 190, 485-96.

EPSTEIN, L. H., J. B. RICHARDS, F. G. SAAD, R. A. PALUCH, J. N. ROEMMICH, AND C. LERMAN (2003): "Comparison between Two Measures of Delay Discounting in Smokers," *Experimental and Clinical Psychopharmacology*, 11, 131-8.

ERIKSEN, M., J. MACKAY, N. SCHLUGER, F. I. GOMESHTAPEH, AND J. DROPE (2015): *The Tobacco Atlas*. Atlanta, GA: American Cancer Society.

FECHNER, G. T. (1966/1860): *Elements of Psychophysics*. New York, NY: Holt, Rinehart and Winston.

FIELD, M., M. SANTARCANGELO, H. SUMNALL, A. GOUDIE, AND J. COLE (2006): "Delay Discounting and the Behavioural Economics of Cigarette Purchases in Smokers: The Effects of Nicotine Deprivation," *Psychopharmacology*, 186, 255-63.

FUDENBERG, D., AND D. K. LEVINE (2006): "A Dual-Self Model of Impulse Control," *American Economic Review*, 96, 1449-1476.

FUDENBERG, D., AND D. K. LEVINE (2011): "Risk, Delay and Convex Self-Control Costs," *American Economic Journal: Microeconomics*, 3, 34-68.

FUDENBERG, D., AND D. K. LEVINE (2012): "Timing and Self-Control," *Econometrica*, 80, 1-42.

GORBER, S. C., S. SCHOFIELD-HURWITZ, J. HARDT, G. LEVASSEUR, AND M. TREMBLAY (2009): "The Accuracy of Self-Reported Smoking: A Systematic Review of the Relationship between Self-Reported and Cotinine-Assessed Smoking Status," *Nicotine & Tobacco Research*, 11, 12-24.

GRUBER, J., AND B. KÖSZEGI (2001): "Is Addiction "Rational"? Theory and Evidence," *Quarterly Journal of Economics*, 116, 1261 - 1303.

GUL, F., AND W. PESENDORFER (2007): "Harmful Addiction," *The Review of Economic Studies*, 74, 147-172.

HARRISON, G. W., AND M. I. LAU (2014): "Risk Attitudes, Sample Selection and Attrition in a Longitudinal Field Experiment," *Working Paper* 2014-04, Center for the Economic Analysis of Risk (CEAR), Georgia State University.

HARRISON, G. W., M. I. LAU, AND E. E. RUTSTRÖM (2009): "Risk Attitudes, Randomization to Treatment, and Self-Selection into Experiments," *Journal of Economic Behavior and Organization*, 70, 498-507.

HARRISON, G. W., M. I. LAU, AND E. E. RUTSTRÖM (2010): "Individual Discount Rates and Smoking: Evidence from a Field Experiment in Denmark," *Journal of Health Economics*, 29, 708-717.

HARRISON, G. W., AND E. E. RUTSTRÖM (2009): "Expected Utility Theory and Prospect Theory: One Wedding and a Decent Funeral," *Experimental Economics*, 12, 133-158.

HARRISON, G. W., AND J. T. SWARTHOUT (2014): "Experimental Payment Protocols and the Bipolar Behaviorist," *Theory and Decision*, 77, 423-438.

HEATHERTON, T. F., L. T. KOZLOWSKI, R. C. FRECKER, AND K. FAGERSTRÖM (1991): "The Fagerström Test for Nicotine Dependence: A Revision of the Fagerström Tolerance Questionnaire," *British Journal of Addiction*, 86, 1119-1127.

HEY, J. D., AND C. ORME (1994): "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62, 1291-1326.

HEYMAN, G. M. (2009): *Addiction: A Disorder of Choice*. Cambridge, MA: Harvard University Press.

HEYMAN, G. M., AND S. P. GIBB (2006): "Delay Discounting in College Cigarette Chippers," *Behavioural Pharmacology*, 17, 669-79.

HOFMEYR, A., J. MONTEROSSO, A. C. DEAN, A. M. MORALES, R. M. BILDER, F. W. SABB, AND E. D. LONDON (2017): "Mixture Models of Delay Discounting and Smoking Behavior," *The American Journal of Drug and Alcohol Abuse*, 43, 271-280.

JOHNSON, M. W., W. K. BICKEL, AND F. BAKER (2007): "Moderate Drug Use and Delay Discounting: A Comparison of Heavy, Light, and Never Smokers," *Experimental and Clinical Psychopharmacology*, 15, 187-94.

JONES, B. A., R. D. LANDES, R. YI, AND W. K. BICKEL (2009): "Temporal Horizon: Modulation by Smoking Status and Gender," *Drug and Alcohol Dependence*, 104 Supplement 1, S87-93.

JUSOT, F., AND M. KHLAT (2013): "The Role of Time and Risk Preferences in Smoking Inequalities: A Population-Based Study," *Addictive Behaviors*, 38, 2167-73.

KANG, M. I., AND S. IKEDA (2014): "Time Discounting and Smoking Behavior: Evidence from a Panel Survey," *Health Economics*, 23, 1443-1464.

KOBIELLA, A., S. RIPKE, N. B. KROEMER, C. VOLLMERT, S. VOLLSTADT-KLEIN, D. E. ULSHOFER, AND M. N. SMOLKA (2014): "Acute and Chronic Nicotine Effects on Behaviour and Brain Activation During Intertemporal Decision Making," *Addiction Biology*, 19, 918-930.

LAIBSON, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112, 443-477.

LAIBSON, D. (2001): "A Cue-Theory of Consumption," *Quarterly Journal of Economics*, 116, 81-119.

LAWYER, S. R., F. SCHOEPFLIN, R. GREEN, AND C. JENKS (2011): "Discounting of Hypothetical and Potentially Real Outcomes in Nicotine-Dependent and Nondependent Samples," *Experimental and Clinical Psychopharmacology*, 19, 263-74.

LOEWENSTEIN, G., T. O'DONOGHUE, AND M. RABIN (2003): "Projection Bias in Predicting Future Utility," *Quarterly Journal of Economics*, 118, 1209-1248.

LOOMES, G., AND R. SUGDEN (1998): "Testing Different Stochastic Specifications of Risky Choice," *Economica*, 65, 581-598.

MAZUR, J. E. (1984): "Tests of an Equivalence Rule for Fixed and Variable Reinforcer Delays," *Journal of Experimental Psychology: Animal Behavior Processes*, 10, 426-437.

MCLACHLAN, G., AND D. PEEL (2000): *Finite Mixture Models*. New York, NY: Wiley.

MELANKO, S., K. LERAAS, C. COLLINS, S. FIELDS, AND B. REYNOLDS (2009): "Characteristics of Psychopathy in Adolescent Nonsmokers and Smokers: Relations to Delay Discounting and Self Reported Impulsivity," *Experimental and Clinical Psychopharmacology*, 17, 258-65.

MITCHELL, S. H. (1999): "Measures of Impulsivity in Cigarette Smokers and Non-Smokers," *Psychopharmacology*, 146, 455-64.

MITCHELL, S. H., AND V. B. WILSON (2012): "Differences in Delay Discounting between Smokers and Nonsmokers Remain When Both Rewards Are Delayed," *Psychopharmacology*, 219, 549-62.

MYERSON, J., L. GREEN, AND M. WARUSAWITHARANA (2001): "Area under the Curve as a Measure of Discounting," *Journal of the Experimental Analysis of Behavior*, 76, 235-43.

ODUM, A. L., AND A. A. BAUMANN (2007): "Cigarette Smokers Show Steeper Discounting of Both Food and Cigarettes Than Money," *Drug and Alcohol Dependence*, 91, 293-6.

ODUM, A. L., G. J. MADDEN, AND W. K. BICKEL (2002): "Discounting of Delayed Health Gains and Losses by Current, Never- and Ex-Smokers of Cigarettes," *Nicotine & Tobacco Research*, 4, 295-303.

OHMURA, Y., T. TAKAHASHI, AND N. KITAMURA (2005): "Discounting Delayed and Probabilistic Monetary Gains and Losses by Smokers of Cigarettes," *Psychopharmacology*, 182, 508-15.

ORPHANIDES, A., AND D. ZERVOS (1995): "Rational Addiction with Learning and Regret," *Journal of Political Economy*, 103, 739-758.

POLTAVSKI, D. V., AND J. N. WEATHERLY (2013): "Delay and Probability Discounting of Multiple Commodities in Smokers and Never-Smokers Using Multiple-Choice Tasks," *Behavioural Pharmacology*, 24, 659-67.

PRATT, J. W. (1964): "Risk Aversion in the Small and in the Large," *Econometrica*, 32, 122-136.

PRELEC, D. (1998): "The Probability Weighting Function," *Econometrica*, 66, 497-527.

QUIGGIN, J. (1982): "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization*, 3, 323-343.

REDISH, A. D., S. JENSEN, AND A. JOHNSON (2008): "A Unified Framework for Addiction: Vulnerabilities in the Decision Process," *Behavioral and Brain Sciences*, 31, 415-487.

REYNOLDS, B. (2004): "Do High Rates of Cigarette Consumption Increase Delay Discounting? A Cross-Sectional Comparison of Adolescent Smokers and Young-Adult Smokers and Nonsmokers," *Behavioural Processes*, 67, 545-9.

REYNOLDS, B. (2006): "The Experiential Discounting Task Is Sensitive to Cigarette-Smoking Status and Correlates with a Measure of Delay Discounting," *Behavioural Pharmacology*, 17, 133-42.

REYNOLDS, B., AND S. FIELDS (2012): "Delay Discounting by Adolescents Experimenting with Cigarette Smoking," *Addiction*, 107, 417-24.

REYNOLDS, B., K. KARRAKER, K. HORN, AND J. B. RICHARDS (2003): "Delay and Probability Discounting as Related to Different Stages of Adolescent Smoking and Non-Smoking," *Behavioural Processes*, 64, 333-344.

REYNOLDS, B., M. PATAK, P. SHROFF, R. B. PENFOLD, S. MELANKO, AND A. M. DUHIG (2007): "Laboratory and Self-Report Assessments of Impulsive Behavior in Adolescent Daily Smokers and Nonsmokers," *Experimental and Clinical Psychopharmacology*, 15, 264-71.

REYNOLDS, B., J. B. RICHARDS, K. HORN, AND K. KARRAKER (2004): "Delay Discounting and Probability Discounting as Related to Cigarette Smoking Status in Adults," *Behavioural Processes*, 65, 35-42.

ROSS, D. (2010): "Economic Models of Pathological Gambling," in *What Is Addiction?*, ed. by D. Ross, H. Kincaid, D. Spurrett, and P. Collins. Cambridge, MA: MIT Press, 131-158.

ROSS, D. (2011): "Estranged Parents and a Schizophrenic Child: Choice in Economics, Psychology and Neuroeconomics," *Journal of Economic Methodology*, 18, 215-229.

ROSS, D. (2014a): *Philosophy of Economics*. Basingstoke: Palgrave Macmillan.

ROSS, D. (2014b): "Psychological Versus Economic Models of Bounded Rationality," *Journal of Economic Methodology*, 21, 411-427.

ROSS, D., C. SHARP, R. VUCHINICH, AND D. SPURRETT (2008): *Midbrain Mutiny: The Behavioural Economics and Neuroeconomics of Disordered Gambling*. Cambridge, MA: MIT Press.

RUBIN, D. B. (1998): "What Does It Mean to Estimate the Causal Effects of 'Smoking'?," *Proceedings of the Section on Statistics in Epidemiology of the American Statistical Association*, 18-27.

RUBIN, D. B. (2001): "Estimating the Causal Effects of Smoking," *Statistics in Medicine*, 20, 1395-1414.

SAHA, A. (1993): "Expo-Power Utility: A 'Flexible' Form for Absolute and Relative Risk Aversion," *American Journal of Agricultural Economics*, 75, 905-913.

SCHELLING, T. C. (1984): *Choice and Consequence*. Cambridge, MA: Harvard Univerity Press.

SHEFFER, C. E., M. MENNEMEIER, R. D. LANDES, W. K. BICKEL, S. BRACKMAN, J. DORNHOFFER, T. KIMBRELL, AND G. BROWN (2013): "Neuromodulation of Delay Discounting, the Reflection Effect, and Cigarette Consumption," *Journal of Substance Abuse Treatment*, 45, 206-14.

SHISANA, O., D. LABADARIOS, T. REHLE, L. SIMBAYI, K. ZUMA, A. DHANSAY, P. REDDY, W. PARKER, E. HOOSAIN, P. NAIDOO, C. HONGORO, Z. MCHIZA, N. P. STEYN, N. DWANE, M. MAKOAE, T. MALULEKE, S. RAMLAGAN, S. ZUNGU, M. G. EVANS, L. JACOBS, M. FABER, AND SANHANES-1 TEAM (2013): *South African National Health and Nutrition Examination Survey (SANHANES-1)*. Cape Town: HSRC Press.

STILLWELL, D. J., AND R. J. TUNNEY (2012): "Effects of Measurement Methods on the Relationship between Smoking and Delay Reward Discounting," *Addiction*, 107, 1003-12.

SWEITZER, M. M., E. C. DONNY, L. C. DIERKER, J. D. FLORY, AND S. B. MANUCK (2008): "Delay Discounting and Smoking: Association with the Fagerstrom Test for Nicotine Dependence but Not Cigarettes Smoked Per Day," *Nicotine & Tobacco Research*, 10, 1571-5.

SZREK, H., L. W. CHAO, S. RAMLAGAN, AND K. PELTZER (2012): "Predicting (Un)Healthy Behavior: A Comparison of Risk-Taking Propensity Measures," *Judgment and Decision Making*, 7, 716-727.

THALER, R. H., AND H. M. SHEFRIN (1981): "An Economic Theory of Self-Control," *Journal of Political Economy*, 89, 392-406.

TVERSKY, A., AND D. KAHNEMAN (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 297-323.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES (2014): *The Health Consequences of Smoking: 50 Years of Progress. A Report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

VAN DER POL, M., AND M. RUGGERI (2008): "Is Risk Attitude Outcome Specific within the Health Domain?," *Journal of Health Economics*, 27, 706-17.

WEST, R. (2006): *Theory of Addiction*. Oxford: Blackwell Publishing.

WILCOX, N. T. (2011): "'Stochastically More Risk Averse:' a Contextual Theory of Stochastic Discrete Choice under Risk," *Journal of Econometrics*, 162, 89-104.

WING, V. C., T. G. MOSS, R. A. RABIN, AND T. P. GEORGE (2012): "Effects of Cigarette Smoking Status on Delay Discounting in Schizophrenia and Healthy Controls," *Addictive Behaviors*, 37, 67-72.

WINSTON, G. C. (1980): "Addiction and Backsliding: A Theory of Compulsive Consumption," *Journal of Economic Behavior and Organization*, 1, 295-324.

WONG, S. L., M. SHIELDS, S. LEATHERDALE, E. MALAISON, AND D. HAMMOND (2012): "Assessment of Validity of Self-Reported Smoking Status," *Health Reports*, 23, 47-53.

YAARI, M. E. (1987): "The Dual Theory of Choice under Risk," *Econometrica*, 55, 95-115.

YI, R., W. D. CHASE, AND W. K. BICKEL (2007): "Probability Discounting among Cigarette Smokers and Nonsmokers: Molecular Analysis Discerns Group Differences," *Behavioural Pharmacology*, 18, 633-9.

TABLE 1: REVIEW OF EXPERIMENTAL LITERATURE ON SMOKING AND DISCOUNTING BEHAVIOUR

Study	Sample (size)	Elicitation method	Task-related incentives (max LL)	Horizon	Front end delay (FED)	Correct for non-linear utility	Models (estimated rates)	Statistical method	Hyperbolicky discounting?	Significant relationship with smoking?
Bickel, Odum and Madden (1999)	Adults in Burlington, VT, USA ($N_S = 23$, $N_{NS} = 22$, $N_{ES} = 21$)	Choice (ordered)	No ($1000)	7 - 9131 days	No	No	H and E ($\delta^H_S = 0.054$) ($\delta^H_{NS} = 0.007$) ($\delta^H_{ES} = 0.007$)	NLLS for discounting, ANOVA and non-parametric tests for analysis	Yes (compared to E) based on R^2 comparisons	Yes, positive for S relative to NS ($p<0.01$) and ES ($p<0.01$); No for NS relative to ES.
Mitchell (1999)	Adults in Durham, NH, USA ($N_S = 20$, $N_{NS} = 20$)	Choice (random)	Yes ($10)	0 - 365 days	No	No	H ($\delta_S = 0.012$) ($\delta_{NS} = 0.006$)	NLLS for discounting, non-parametric tests for analysis	By assumption	Yes ($p<0.06$), positive.
Baker, Johnson and Bickel (2003)	Adults in Burlington, VT, USA ($N_S = 30$, $N_{NS} = 30$)	Titration (random - Richards et al. (1999))	Yes ($100)	Real: 1 - 183 days	No	No	H NRD but from Figure 2: ($10: $\delta_S = 0.008$, $\delta_{NS} = 0.001$) ($100: $\delta_S = 0.005$, $\delta_{NS} = 0.001$)	NLLS for discounting, ANOVA for analysis	By assumption	Real: Yes ($p<0.01$), positive.
			No ($1000)	Hypothetical: 1 - 9131 days	No	No	H NRD but from Figure 2: ($10: $\delta_S = 0.008$, $\delta_{NS} = 0.003$) ($100: $\delta_S = 0.006$, $\delta_{NS} = 0.0008$) ($1000: $\delta_S = 0.004$, $\delta_{NS} = 0.0005$)			Hypothetical: Yes ($p<0.01$), positive.
Reynolds, Karraker, Horn and Richards (2003)	Adolescents in Morgantown, WV, USA ($N_S = 19$, $N_{NS} = 19$, $N_T = 17$)	Titration (random - Richards et al. (1999))	1-out-of-2-tasks ($10)	1 - 365 days	No	No	H ($\delta_S = 0.010$) ($\delta_{NS} = 0.007$) ($\delta_T = 0.016$)	NLLS for discounting, ANOVA for analysis	By assumption	No.
Reynolds (2004)	Adolescents and young adults in Morgantown, WV, USA ($N_{S(adolescent)} = 19$, $N_{S(adult)} = 25$, $N_{NS} = 29$)	Titration (random - Richards et al. (1999))	1-out-of-2-tasks ($10)	1 - 365 days	No	No	H ($\delta_{S(adolescent)} = 0.016$) ($\delta_{S(adult)} = 0.075$) ($\delta_{NS(adult)} = 0.012$)	NLLS for discounting, ANOVA, correlations and post hoc tests for analysis	By assumption	Yes, positive for $S_{(adult)}$ relative to $S_{(adolescent)}$ ($p<0.05$) and $NS_{(adult)}$ ($p<0.05$). No for $S_{(adolescent)}$ relative to $NS_{(adult)}$.
Reynolds, Richards, Horn and Karraker (2004)	Mostly students in Morgantown, WV, USA ($N_S = 25$, $N_{NS} = 29$)	Titration (random - Richards et al. (1999))	1-out-of-2-tasks ($10)	1 - 365 days	No	No	H ($\delta_S = 0.066$) ($\delta_{NS} = 0.015$)	NLLS for discounting, ANOVA for analysis	By assumption	Yes ($p<0.05$), positive.
Ohmura, Takahashi and Kitamura (2005)	Students in Sapporo, Japan ($N_S = 27$, $N_{NS} = 23$)	Titration (random - Richards et al. (1999))	No (¥100,000 = $1000)	7 - 1826 days)	No	No	H, E and AUC. ($AUC_S = 0.54$) ($AUC_{NS} = 0.58$)	AUC and NLLS for discounting, correlations and t tests for analysis	Yes (compared to E) based on R^2 comparisons	No.

Source: Authors' construction. See Appendix A for more details on these studies.

Notes: S = smoker; NS = non-smoker/never-smoker; ES = ex-smoker; LS = light smoker; T = trier; FS = fast smoking adopter; SS = slow smoking progressor.

H = hyperbolic; E = exponential; QH = quasi-hyperbolic; AUC = area under the curve; NRD = not reported directly; [a] = annual rate; [b] = weekly rate; NLLS = non-linear least squares; ML = maximum likelihood.

ANOVA = analysis of variance; ANCOVA = analysis of covariance

TABLE 1: REVIEW OF EXPERIMENTAL LITERATURE ON SMOKING AND DISCOUNTING BEHAVIOUR (CONTINUED)

Study	Sample (size)	Elicitation method	Task-related incentives (max LL)	Horizon	Front end delay (FED)	Correct for non-linear utility	Models (estimated rates)	Statistical method	Hyperbolicky discounting?	Significant relationship with smoking?
Heyman and Gibb (2006)	Students in Cambridge, MA, USA ($N_S = 19$, $N_{NS} = 31$, $N_{LS} = 21$)	Choice (ordered)	Yes ($29)	Real: 1 – 30 days	No	No	H Real: ($\delta_S = 0.074$) ($\delta_{NS} = 0.036$) ($\delta_{LS} = 0.045$)	Algebra and averaging for discounting, F-test and post-hoc tests for analysis	By assumption	Real: Yes, positive for S relative to NS ($p<0.01$) and LS ($p<0.05$); No for LS relative to NS.
			No ($1000)	Hypothetical: 7 – 3650 days	No	No	H Hypothetical: ($\delta_S = 0.007$) ($\delta_{NS} = 0.009$) ($\delta_{LS} = 0.004$)			Hypothetical: No.
Reynolds (2006)	Adults in Buffalo, NY, USA ($N_S = 15$, $N_{NS} = 15$)	Titration (random – Richards et al. (1999))	No ($10)	1 – 365 days	No	No	H ($\delta_S = 0.088$) ($\delta_{NS} = 0.020$)	NLLS for discounting, non-parametric tests for analysis	By assumption	Yes ($p<0.01$), positive.
Johnson, Bickel and Baker (2007)	Adults in Burlington, VT, USA ($N_S = 30$, $N_{NS} = 30$, $N_{LS} = 30$)	Titration (random – Richards et al. (1999))	Yes ($100)	Real: 1 – 183 days	No	No	H Real: ($10: $\delta_S = 0.006$, $\delta_{LS} = 0.003$, $\delta_{NS} = 0.0009$) ($100: $\delta_S = 0.003$, $\delta_{LS} = 0.001$, $\delta_{NS} = 0.0008$)	NLLS for discounting, ANOVA for analysis	By assumption	Real: Yes, positive for S ($p<0.05$) and LS ($p<0.05$) relative to NS; No for S relative to LS.
			No ($1000)	Hypothetical: 1 – 9131 days	No	No	H Hypothetical: ($10: $\delta_S = 0.006$, $\delta_{LS} = 0.007$, $\delta_{NS} = 0.002$) ($100: $\delta_S = 0.004$, $\delta_{LS} = 0.002$, $\delta_{NS} = 0.0005$) ($1000: $\delta_S = 0.002$, $\delta_{LS} = 0.0008$, $\delta_{NS} = 0.0003$)			Hypothetical: Yes, positive for S ($p<0.01$) and LS ($p<0.05$) relative to NS; No for S relative to LS.
Reynolds et al. (2007)	Adolescents in Columbus, OH, USA ($N_S = 25$, $N_{NS} = 26$)	Titration (random – Richards et al. (1999))	Yes ($10)	1 – 365 days	No	No	AUC NRD but from Figure 1: ($AUC_S = 0.129$) $AUC_{NS} = 0.234$)	AUC for discounting, ANOVA and ANCOVA for analysis	AUC, but dropped subjects that had poor H fit	Yes ($p<0.05$), positive.
Bickel, Yi, Kowal and Gatchalian (2008)	Adults in Little Rock, AR, USA ($N_S = 30$, $N_{NS} = 29$)	Titration (random - Richards et al. (1999))	No ($1000)	1 - 9131 days	No	No	H and E ($\delta^H_S = 0.007$) ($\delta^H_{NS} = 0.001$)	NLLS for discounting, ANCOVA for analysis	Yes (compared to E) based on R^2 comparisons	Yes ($p<0.05$), positive.

Source: Authors' construction. See Appendix A for more details on these studies.

Notes: S = smoker; NS = non-smoker/never-smoker; ES = ex-smoker; LS = light smoker; T = trier; FS = fast smoking adopter; SS = slow smoking progressor.

H = hyperbolic; E = exponential; QH = quasi-hyperbolic; AUC = area under the curve; NRD = not reported directly; [a] = annual rate; [b] = weekly rate; NLLS = non-linear least squares; ML = maximum likelihood.

ANOVA = analysis of variance; ANCOVA = analysis of covariance

TABLE 1: REVIEW OF EXPERIMENTAL LITERATURE ON SMOKING AND DISCOUNTING BEHAVIOUR (CONTINUED)

Study	Sample (size)	Elicitation method	Task-related incentives (max LL)	Horizon	Front end delay (FED)	Correct for non-linear utility	Models (estimated rates)	Statistical method	Hyperbolicky discounting?	Significant relationship with smoking?
Chabris et al. (2008)	1. Adults in Boston, MA, USA (N = 126)	Choice (random – Kirby et al. (1999))	1-in-6-chance ($85)	7 – 186 days	No	No	H (δ = 0.015, SD = 0.02)	ML for discounting, OLS, Tobit, Probit for analysis	By assumption	Yes ($p<0.05$), positive.
	2. Adults in the USA (recruited online) (N = 326)		1-in-6-chance ($85)	7 – 186 days	No	No	H (δ = 0.008, SD = 0.009)	ML for discounting, OLS, Tobit, Probit for analysis	By assumption	No
Sweitzer et al. (2008)	Adults in Allegheny County, PA, USA (N_S = 101, N_{NS} = 145, N_T = 279, N_{ES} = 185)	Choice (random)	No ($100)	7 - 1825 days	No	No	H (δ_S = 0.120) (δ_{NS} = 0.079) (δ_T = 0.090) (δ_{ES} = 0.086)	NLLS for discounting, ANCOVA for analysis	By assumption	Yes, positive for S relative to NS ($p<0.01$), ES ($p<0.01$) and T ($p<0.01$); No for all other comparisons.
Adams and Nettle (2009)	Adults in 15 major urban areas in the USA (recruited online) (N_S = 70, N_{NS} = 346)	Choice (ordered)	No ($1000)	30 - 3652 days	No	No	H (δ = 1.3)[a]	NLLS for discounting, logistic regression for analysis	By assumption	No.
Audrain-McGovern et al. (2009)	High school students in northern Virginia, USA (N_{NS} = 556, N_{FS} = 112, N_{SS} = 241)	Choice (random - Kirby et al. (1999))	Not reported ($85)	7 - 186 days	No	No	H Assuming ln transformation: (δ_{FS} = 0.023) (δ_{SS} = 0.016) (δ_{NS} = 0.010)	Algebra and averaging for discounting, latent growth curve modeling (LGCM) and growth mixture modeling (GMM) for analysis	By assumption	LCGM: Yes ($p<0.05$), positive. GMM: Yes, positive for FS ($p<0.05$) and SS ($p<0.05$) relative to NS; No for FS relative to SS.
Jones, Landes, Yi and Bickel (2009)	Adults in Little Rock, AR, USA (N_S = 86, N_{NS} = 141)	Titration (ordered or random)	No ($1000)	1 - 9131 days	No	No	H NRD but from Figure 3: ($100: $\delta_{S(men)}$ = 0.012, $\delta_{NS(men)}$ = 0.001, $\delta_{S(women)}$ = 0.0015, $\delta_{NS(women)}$ =0.002) ($1000: $\delta_{S(men)}$ = 0.0075, $\delta_{NS(men)}$ = 0.0005, $\delta_{S(women)}$ = 0.001, $\delta_{NS(women)}$ =0.001)	NLLS for discounting, ANCOVA for analysis	By assumption	Yes, positive for $S_{(men)}$ ($p<0.01$) relative $NS_{(men)}$ at $100 and $1000 magnitudes; No for $S_{(women)}$ relative to $NS_{(women)}$ at both magnitudes.
Melanko et al. (2009)	Adolescents in central Ohio, USA (N_S = 50, N_{NS} = 25). Smokers were split into high and low pyschopathology groups.	Titration (random - Richards et al. (1999))	Yes ($10)	1 - 365 days	No	No	AUC NRD but from Figure 1: ($AUC_{S(low)}$ = 0.126) ($AUC_{S(high)}$ = 0.214) (AUC_{NS} = 0.275)	AUC for discounting, ANOVA for analysis	AUC, no assumption about form of discounting	Yes, positive for $S_{(low)}$ relative to NS ($p=0.01$); No for all other comparisons.
Businelle, McVay, Kendzor and Copeland (2010)	Adults in southern USA (N_S = 20, N_{NS} = 34)	Choice (ordered)	No ($1000)	0.25 - 9131 days	No	No	H and AUC (δ_S = 0.077) (δ_{NS} = 0.039)	NLLS and AUC for discounting, ANCOVA for analysis	By assumption (but also used AUC)	Yes ($p=0.01$), positive.

Source: Authors' construction. See Appendix A for more details on these studies.

Notes: S = smoker; NS = non-smoker/never-smoker; ES = ex-smoker; LS = light smoker; T = trier; FS = fast smoking adopter; SS = slow smoking progressor.

H = hyperbolic; E = exponential; QH = quasi-hyperbolic; AUC = area under the curve; NRD = not reported directly; [a] = annual rate; [b] = weekly rate; NLLS = non-linear least squares; ML = maximum likelihood.

ANOVA = analysis of variance; ANCOVA = analysis of covariance

TABLE 1: REVIEW OF EXPERIMENTAL LITERATURE ON SMOKING AND DISCOUNTING BEHAVIOUR (CONTINUED)

Study	Sample (size)	Elicitation method	Task-related incentives (max LL)	Horizon	Front end delay (FED)	Correct for non-linear utility	Models (estimated rates)	Statistical method	Hyperbolicky discounting?	Significant relationship with smoking?
Harrison, Lau and Rutström (2010)	Adults in Denmark (N_S = 71, N_{NS} = 181)	Choice (ordered)	1-in-10-chance ($1175)	30 - 730 days	Yes	Yes	H and E Linear utility: ($\delta^H_{S(men)} = 0.341)^a$ ($\delta^H_{NS(men)} = 0.240)^a$ ($\delta^H_{S(women)} = 0.329)^a$ ($\delta^H_{NS(women)} = 0.250)^a$	ML for discounting and analysis	25% - 40% of choices by smokers and non-smokers best characterised by H	Linear utility: Men: Yes (p<0.05), positive; Women: Yes (p<0.10), positive.
							H and E Concave utility: ($\delta^H_{S(men)} = 0.127)^a$ ($\delta^H_{NS(men)} = 0.093)^a$ ($\delta^H_{S(women)} = 0.109)^a$ $\delta^H_{NS(women)} = 0.095)^a$			Concave utility: Men: Yes (p<0.05), positive; Women: No.
Bickel et al. (2012)	Adults in the USA (recruited online) (N_S = 182, N_{NS} = 614)	Choice (random)	No ($85)	10 - 75 days	No	No	H (Not reported)	Algebra and averaging for discounting, ANCOVA for analysis	By assumption	Yes (p<0.01), positive.
Mitchell and Wilson (2012)	1. Adults in Portland, OR, USA (N_S = 20, N_{NS} = 20)	Choice (random)	Yes ($50)	14 - 154 days	Yes	No	H and QH (0 FED: $\delta^H_S = 0.230$, $\delta^H_{NS} = 0.020$) (+ FED: $\delta^H_S = 0.070$, $\delta^H_{NS} = 0.010$)	NLLS and ML for discounting, ANOVA for analysis	By assumption (but also estimated QH)	Yes (p<0.01), positive.
	2. Adults in Portland, OR, USA (N_S = 16, N_{NS} = 16)		No ($50)	14 - 154 days	Yes	No	H and QH (0 FED: $\delta^H_S = 0.120$, $\delta^H_{NS} = 0.020$) (+ FED: $\delta^H_S = 0.050$, $\delta^H_{NS} = 0.010$)			Yes (p<0.01), positive.
Reynolds and Fields (2012)	Adolescents in Columbus, OH, USA (N_S = 50, N_{NS} = 50, N_T = 41)	Titration (random - Richards et al. (1999))	Yes ($10)	1 - 365 days	No	No	AUC NRD but from Figure 1: (AUC_S = 0.166) (AUC_T = 0.224) (AUC_{NS} = 0.347)	AUC for discounting, ANOVA and ANCOVA for analysis	AUC, no assumption about form of discounting	Yes, positive for S (p<0.01) and T (p<0.05) relative to NS; No for S relative to T.
Stillwell and Tunney (2012)	International online study (N_S = 1592, N_{LS} = 669, N_{NS} = 6777)	Choice (ordered or random)	No ($1000)	7 - 1826 days	No	No	H NRD but from Figure 3: ($\delta_S = 0.437$) ($\delta_{LS} = 0.397$) ($\delta_{NS} = 0.369$)	NLLS for discounting, ANOVA for analysis	Yes (compared to E) based RSS comparisons	Yes, positive for S relative to LS (p<0.01) and NS (p<0.01) and positive for LS relative to NS (p<0.01).
Wing, Moss, Rabin and George (2012)	Adults in the greater Toronto area, Canada (N_S = 23, N_{NS} = 37)	Choice (random - Kirby et al. (1999))	No ($85)	7 - 186 days	No	No	H NRD but from Figure 1: ($\delta_S = 0.017$) ($\delta_{NS} = 0.011$)	Algebra and averaging for discounting, ANCOVA for analysis	By assumption	No.

Source: Authors' construction. See Appendix A for more details on these studies.

Notes: S = smoker; NS = non-smoker/never-smoker; ES = ex-smoker; LS = light smoker; T = trier; FS = fast smoking adopter; SS = slow smoking progressor.

H = hyperbolic; E = exponential; QH = quasi-hyperbolic; AUC = area under the curve; NRD = not reported directly; [a] = annual rate; [b] = weekly rate; NLLS = non-linear least squares; ML = maximum likelihood.

ANOVA = analysis of variance; ANCOVA = analysis of covariance

TABLE 1: REVIEW OF EXPERIMENTAL LITERATURE ON SMOKING AND DISCOUNTING BEHAVIOUR (CONTINUED)

Study	Sample (size)	Elicitation method	Task-related incentives (max LL)	Horizon	Front end delay (FED)	Correct for non-linear utility	Models (estimated rates)	Statistical method	Hyperbolicky discounting?	Significant relationship with smoking?
Balevich, Wein and Flory (2013)	Students in Flushing, NY, USA ($N_S = 50$, $N_{NS} = 102$, $N_T = 91$)	Choice (random) or titration (random)	No ($100)	1 - 1825 days	No	No	H ($\delta_S = 0.126$) ($\delta_{NS} = 0.135$) ($\delta_T = 0.138$)	NLLS for discounting, ANOVA for analysis	By assumption	No.
Poltavski and Weatherly (2013)	Students in Grand Forks, ND, USA ($N_S = 16$, $N_{LS} = 74$, $N_{NS} = 92$)	Choice (random)	No ($100,000)	183 - 3652 days	No	No	H and AUC ($1000: δ_S= 0.010, δ_{LS} = 0.010, δ_{NS} = 0.007) ($100,000: δ_S = 0.008, δ_{LS} = 0.008, δ_{NS} = 0.007)	NLLS and AUC for discounting, ANOVA for analysis	By assumption (but also used AUC)	No.
Sheffer et al. (2013)	Adults in Little Rock, AR, USA ($N_S = 47$, $N_{NS} = 19$)	Titration (random - Richards et al. (1999))	No ($1000)	1 - 9131 days	No	No	H NRD but from Figure 1: ($\delta_S = 0.020$) ($\delta_{NS} = 0.004$	NLLS for discounting, ANCOVA for analysis	By assumption	Yes ($p<0.05$), positive.
Kang and Ikeda (2014)	Adults in Japan ($N_S \approx 862$, $N_{NS} \approx 2588$)	Choice (ordered)	No (¥1000,000 = $10000)	7 - 365 days	Yes	No	E and proxies for H (See Table III, the mean of δ^E ranges from 0.022 to 1.904)[a]	ML for discounting, hurdle model for analysis	Assumes E but constructs H proxies	Yes ($p<0.01$), positive.
Kobiella et al. (2014)	Adults in Mannheim, Germany ($N_S = 27$, $N_{NS} = 31$)	Choice (random)	Yes (€41.32)	14 - 28 days	Yes	No	H ($\delta_S = 0.055$)[b] ($\delta_{NS} = 0.038$)[b]	NLLS for discounting, t-tests for analysis	By assumption	Yes ($p<0.05$), positive.
Hofmeyr et al. (2017)	Adults in Los Angeles, CA, USA ($N_S = 163$, $N_{NS} = 834$, $N_{ES} = 208$)	Choice (random - Kirby et al. (1999))	UCLA: No ($85) USC: 1-out-of-2-tasks ($85)	7-186 days	No	No	H, E and QH ($\delta^H_S = 0.021$) ($\delta^H_{NS} = 0.012$) ($\delta^H_{ES} = 0.013$)	ML for discounting and analysis	41% - 52% of choices best characterised by H	Yes, positive for S relative to NS ($p<0.01$) and ES ($p<0.01$); No for ES relative to NS.

Source: Authors' construction. See Appendix A for more details on these studies.

Notes: S = smoker; NS = non-smoker/never-smoker; ES = ex-smoker; LS = light smoker; T = trier; FS = fast smoking adopter; SS = slow smoking progressor.

H = hyperbolic; E = exponential; QH = quasi-hyperbolic; AUC = area under the curve; NRD = not reported directly; [a] = annual rate; [b] = weekly rate; NLLS = non-linear least squares; ML = maximum likelihood.

ANOVA = analysis of variance; ANCOVA = analysis of covariance

TABLE 2: REVIEW OF EXPERIMENTAL LITERATURE ON SMOKING AND RISK PREFERENCES

Study	Sample (size)	Elicitation method	Incentives (max prize)	Probabilities	Models (estimated rates)	Statistical method (valid?)	Significant relationship with smoking?
Mitchell (1999)	Adults in Durham, NH, USA ($N_S = 20$, $N_{NS} = 20$)	Choice (random)	Yes ($10)	0.1, 0.25, 0.5, 0.75, 0.9, 1	PD ($\gamma_S = 1.328$) ($\gamma_{NS} = 1.371$)	NLLS for risk aversion, non-parametric tests for analysis. (not valid)	No.
Reynolds, Karraker, Horn and Richards (2003)	Adolescents in Morgantown, WV, USA ($N_S = 19$, $N_{NS} = 19$, $N_T = 17$)	Titration (random - Richards et al. (1999))	1-out-of-2-tasks ($10)	0.25, 0.5, 0.75, 0.9, 1	PD NRD but from Figure 2: ($\gamma_S = 1.610$) ($\gamma_{NS} = 1.110$) ($\gamma_T = 3.820$)	NLLS for risk aversion, ANOVA for analysis. (not valid)	Yes, positive for T relative S ($p<0.05$) and NS ($p<0.05$); No for S relative to NS.
Reynolds, Richards, Horn and Karraker (2004)	Mostly students in Morgantown, WV, USA ($N_S = 25$, $N_{NS} = 29$)	Titration (random - Richards et al. (1999))	1-out-of-2-tasks ($10)	0.25, 0.5, 0.75, 0.9, 1	PD ($\gamma_S = 1.910$) ($\gamma_{NS} = 1.470$)	NLLS for risk aversion, ANOVA for analysis. (not valid)	Yes ($p<0.05$), positive (smokers are more risk averse)
Ohmura, Takahashi and Kitamura (2005)	Students in Sapporo, Japan ($N_S = 27$, $N_{NS} = 23$)	Titration (random - Richards et al. (1999))	No (¥100,000 = $1000)	0.1, 0.3, 0.5, 0.7, 0.9	PD and AUC $AUC_S = 0.230$ $AUC_{NS} = 0.180$	AUC and NLLS for risk aversion, correlations and t-tests for analysis. (not valid)	Yes ($p=0.08$), negative (smokers are less risk averse)
Reynolds (2006)	Adults in Buffalo, NY, USA ($N_S = 15$, $N_{NS} = 15$)	Titration (random - Richards et al. (1999))	No ($10)	0.25, 0.5, 0.75, 0.9, 1	PD ($\gamma_S = 3.908$) ($\gamma_{NS} = 1.574$)	NLLS for risk aversion, non-parametric tests for analysis. (not valid)	Yes ($p<0.05$), positive (smokers are more risk averse)
Reynolds et al. (2007)	Adolescents in Columbus, OH, USA ($N_S = 25$, $N_{NS} = 26$)	Titration (random - Richards et al. (1999))	Yes ($10)	0.25, 0.5, 0.75, 0.9, 1	AUC and PD (Not reported)	AUC for risk aversion, ANOVA for analysis. (not valid)	No.
Yi, Chase and Bickel (2007)	Adults in Little Rock, AR, USA ($N_S = 30$, $N_{NS} = 29$)	Titration (ordered)	No ($1000)	0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 0.95	PD and AUC (Not reported)	NLLS and AUC for risk aversion, ANOVA for analysis. (not valid)	No when analysing all the data; Yes ($p<0.05$), positive, when using only probabilities ≥ 0.5.
Anderson and Mellor (2008)	Adults subjects in Williamsburg, VA, USA ($N_S \approx 79$, $N_{NS} \approx 898$)	Choice (ordered - MPL)	Yes ($11.55)	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1	CRRA (1-r) ($r = 0.257$)	Algebra and averaging for risk aversion, probit model for analysis. (not valid)	Yes ($p<0.1$), negative (smokers are less risk averse).
Harrison, Lau and Rutström (2010)	Adults in Denmark ($N_S = 71$, $N_{NS} = 181$)	Choice (ordered - MPL)	1-in-10-chance ($687)	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1	CRRA (1-r) ($r_{S(men)} = 0.729$) ($r_{NS(men)} = 0.746$) ($r_{S(women)} = 0.811$) ($r_{NS(women)} = 0.755$)	ML for risk aversion and analysis. (valid)	Men: No; Women: Yes ($p<0.06$), positive (smokers are more risk averse)
Szrek, Chao, Ramlagan and Peltzer (2012)	Adults in Witbank, South Africa ($N_S \approx 59$, $N_{NS} \approx 292$)	Choice (ordered - MPL)	Yes (R48 ≈ $7)	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1	CRRA (1-r) ($r = 0.35$, SD = 0.62)	Algebra and averaging for risk aversion, logit model for analysis. (not valid)	No.
Poltavski and Weatherly (2013)	Students in Grand Forks, ND, USA ($N_S = 16$, $N_{LS} = 74$, $N_{NS} = 92$)	Choice (random)	No ($100,000)	0.01, 0.1, 0.5, 0.9, 0.99	PD and AUC ($1000: $\gamma_S = 0.118$, $\gamma_{LS} = 0.134$, $\gamma_{NS} = 0.307$) ($100,000: $\gamma_S = 0.031$, $\gamma_{LS} = 0.167$, $\gamma_{NS} = 0.181$)	NLLS and AUC for risk aversion, ANOVA for analysis. (not valid)	Yes, negative for S relative to NS ($p<0.05$); No for all other comparisons.

Source: Authors' construction. See Appendix B for more details on these studies.

Notes: S = smoker; NS = non-smoker/never-smoker; LS = light smoker; T = trier; PD = probability discounting; AUC = area under the curve; NRD = not reported directly; MPL = multiple price list.
NLLS = non-linear least squares; ML = maximum likelihood; ANOVA = analysis of variance.

TABLE 3
DESCRIPTIVE STATISTICS

Variable	Mean	Std Deviation
Demographics		
Age	19.789	1.815
White	0.417	0.495
Male	0.549	0.499
Commerce faculty	0.674	0.470
Financial aid	0.314	0.466
Smoke	0.617	0.487
Treatments		
Risk task first	0.514	0.501
FED: 0 days	0.343	0.475
FED: 1 week	0.326	0.469
FED: 2 weeks	0.331	0.471
High Principal	0.498	0.500

TABLE 4: RDU THEORY ML ESTIMATES
HETEROGENOUS PREFERENCES

	Model	
	Prelec	
	Estimate	Std Error
Power function parameter (r)		
Age	-0.004	0.011
White	0.029	0.051
Male	0.062	0.049
Commerce faculty	0.030	0.062
Financial aid	-0.051	0.058
Risk task first	-0.015	0.050
Smoker	-0.005	0.055
Constant	0.366	0.230
PWF parameter (φ)		
Age	-0.003	0.006
White	0.001	0.047
Male	-0.009	0.044
Commerce faculty	-0.084	0.120
Financial aid	0.034	0.056
Risk task first	0.054	0.080
Smoker	0.028	0.049
Constant	0.871***	0.206
PWF parameter (η)		
Age	-0.027	0.046
White	-0.062	0.121
Male	-0.166	0.137
Commerce faculty	-0.216	0.184
Financial aid	-0.014	0.139
Risk task first	0.166	0.153
Smoker	0.146	0.177
Constant	1.425**	0.676
Error (μ)		
Constant	0.166***	0.008
N	7000	
log-likelihood	-4119.762	

Results account for clustering at the individual level
* $p<0.10$, ** $p<0.05$, *** $p<0.01$

TABLE 5: NUMBER OF CIGARETTES CONDITIONAL MARGINAL EFFECTS FOR δ

	Model 1	**Model 2**	**Model 3**	**Model 4**
	Exponential	Hyperbolic	Quasi-Hyperbolic	Weibull
Number of cigarettes				
0	0.052 (0.015)	0.044 (0.013)	0.053 (0.014)	0.018 (0.006)
5	0.031 (0.009)	0.025 (0.007)	0.032 (0.009)	0.011 (0.003)
10	0.010 (0.006)	0.006 (0.005)	0.011 (0.005)	0.004 (0.002)
15	-0.011 (0.009)	-0.013 (0.010)	-0.010 (0.008)	-0.003 (0.002)
20	-0.032 (0.015)	-0.032 (0.016)	-0.030 (0.014)	-0.010 (0.004)
25	-0.053 (0.022)	-0.051 (0.023)	-0.051 (0.020)	-0.017 (0.006)

Standard errors in parentheses

TABLE 6: MIXTURE MODEL ML ESTIMATES
RANK-DEPENDENT UTILITY AND HETEROGENOUS PREFERENCES

	Estimate	Std error	*p*-value	95% Confidence Interval	
Power function parameter (r)					
Number of Cigarettes	0.016***	0.005	0.003	0.005	0.026
Number of Cigarettes^2	-0.001***	0.000	0.003	-0.001	0.000
Constant	0.305***	0.028	0.000	0.251	0.359
PWF parameter (φ)					
Number of Cigarettes	-0.013	0.012	0.267	-0.035	0.010
Number of Cigarettes^2	0.001	0.001	0.146	0.000	0.002
Constant	0.813***	0.036	0.000	0.743	0.883
PWF parameter (η)					
Number of Cigarettes	0.018	0.019	0.362	-0.020	0.055
Number of Cigarettes^2	0.000	0.001	0.993	-0.002	0.002
Constant	0.818***	0.047	0.000	0.725	0.910
Discounting parameter (δ_E^{mix})					
Number of Cigarettes	0.031***	0.010	0.002	0.012	0.051
Number of Cigarettes^2	-0.002***	0.001	0.003	-0.003	-0.001
Constant	0.116***	0.019	0.000	0.078	0.153
Discounting parameter (δ_H^{mix})					
Number of Cigarettes	0.047***	0.017	0.005	0.014	0.079
Number of Cigarettes^2	-0.002***	0.001	0.002	-0.003	-0.001
Constant	0.640***	0.076	0.000	0.491	0.790
Mixture probability (π^E)					
Age	-0.002	0.016	0.900	-0.033	0.029
White	0.076	0.086	0.374	-0.092	0.244
Male	-0.105	0.074	0.156	-0.250	0.040
Commerce faculty	-0.017	0.089	0.850	-0.192	0.158
Financial aid	-0.118	0.084	0.160	-0.283	0.047
Risk task first	-0.049	0.080	0.541	-0.205	0.107
FED: 1 week	-0.105	0.083	0.203	-0.267	0.057
FED: 2 weeks	-0.038	0.091	0.679	-0.216	0.141
High Principal	0.186***	0.051	0.000	0.087	0.285
Number of Cigarettes	-0.018***	0.007	0.007	-0.031	-0.005
Constant	0.512	0.313	0.102	-0.101	1.125
Error terms					
Risk error (μ)	0.166***	0.007	0.000	0.151	0.180
Time error (ν)	0.051***	0.012	0.000	0.026	0.075
N	17500				
log-likelihood	-8484.767				

Results account for clustering at the individual level
* *p*<0.10, ** *p*<0.05, *** *p*<0.01

FIGURES

Figure 1

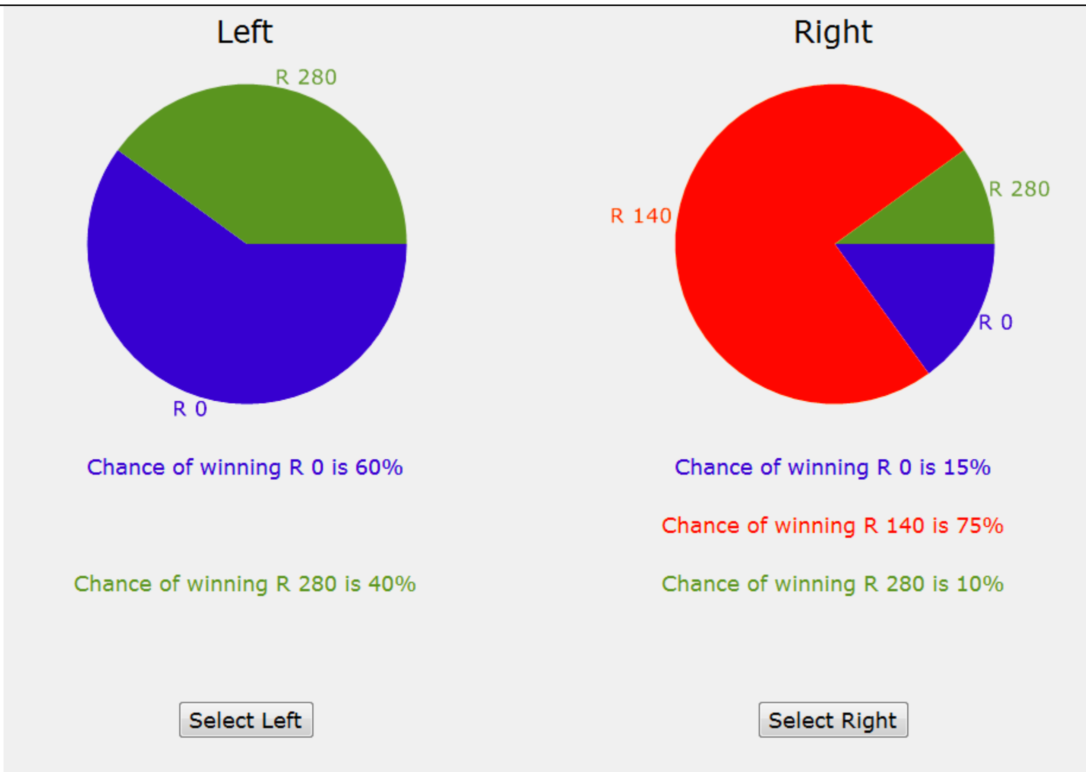


Figure 2: MM triangles of lotteries in the risk preference task

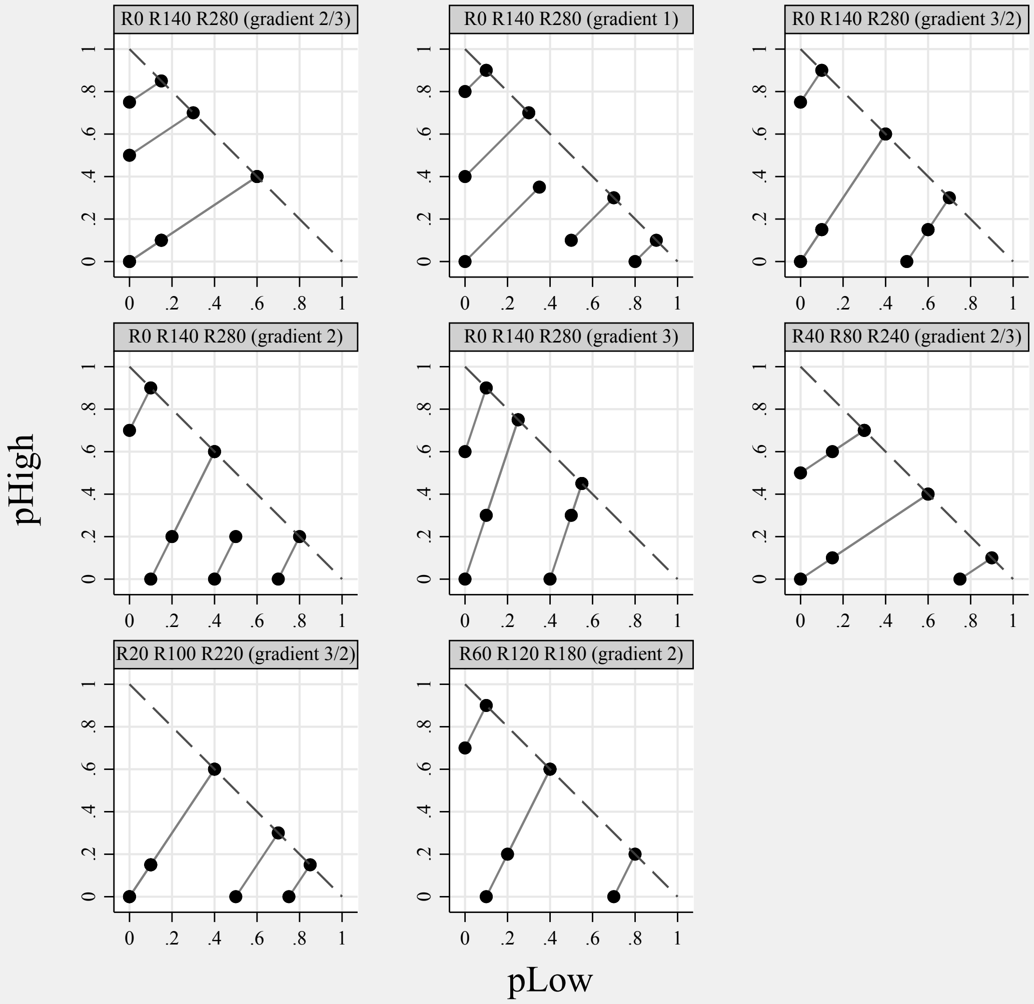


Figure 3

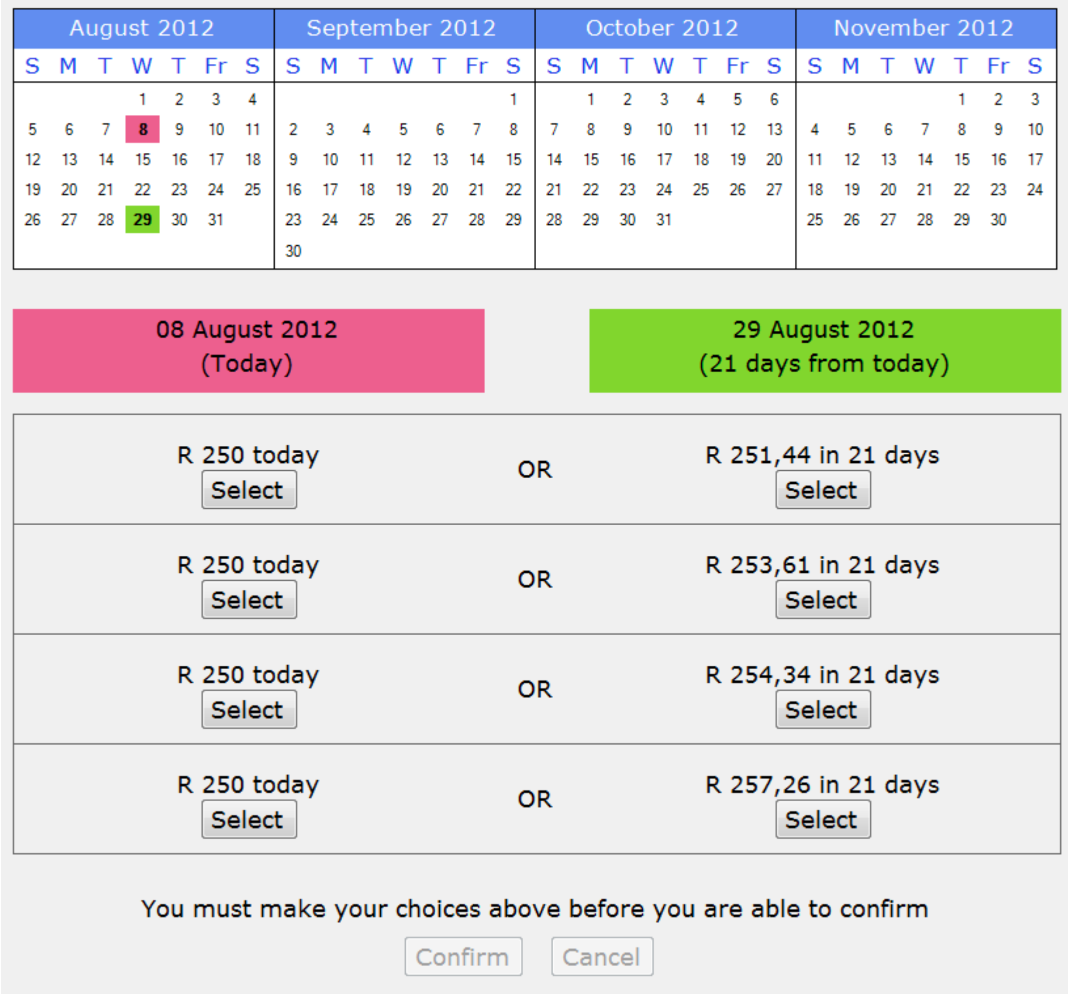
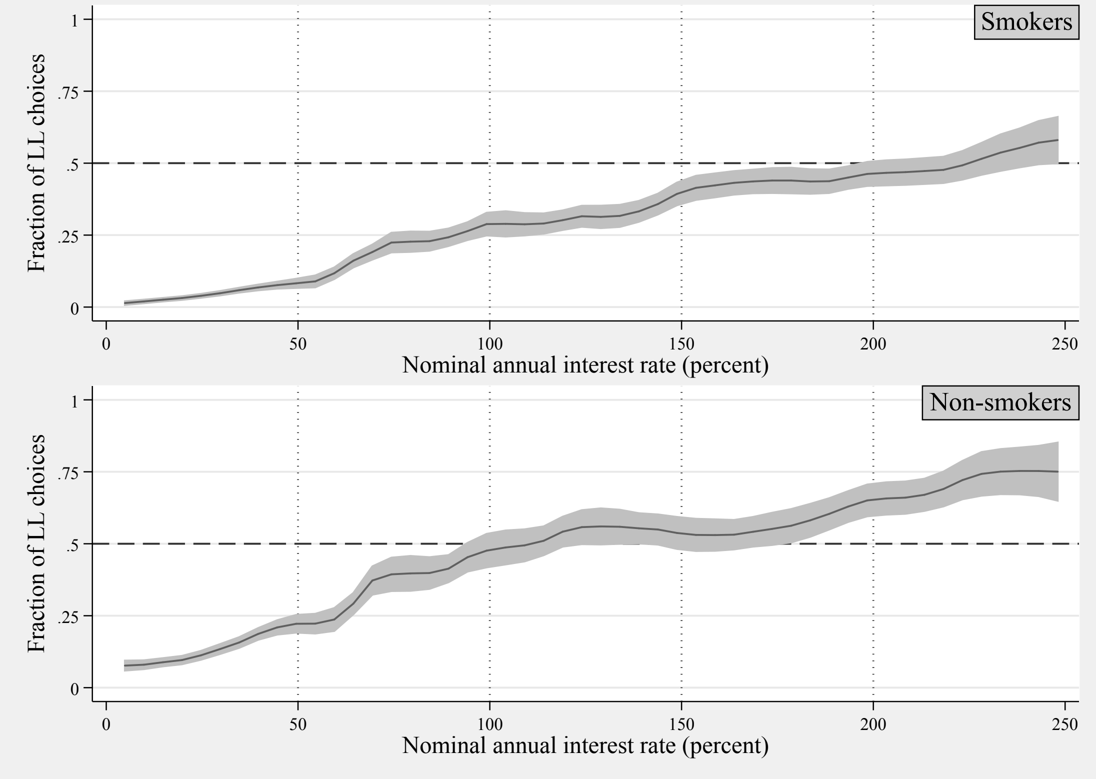

August 2012
September 2012
October 2012
November 2012
08 August 2012
(Today)
29 August 2012
(21 days from today)
R 250 today
Select
OR
R 251,44 in 21 days
Select
R 250 today
Select
OR
R 253,61 in 21 days
Select
R 250 today
Select
OR
R 254,34 in 21 days
Select
R 250 today
Select
OR
R 257,26 in 21 days
Select
You must make your choices above before you are able to confirm
Confirm
Cancel

Figure 4: Fraction of LL choices and interest rate offered



APPENDIX A
[ONLINE WORKING PAPER]

As discussed in the main text, Table 1 provides a detailed summary of experimental studies investigating the relationship between smoking and time preferences. In this appendix we discuss Table 1 in more detail.

Mitchell (1999) and Bickel, Odum and Madden (1999) conducted the first experiments investigating the relationship between smoking and discounting behaviour. Mitchell (1999) presented 20 relatively heavy[1], current smokers (N_S = 20) and 20 never-smokers (N_{NS} = 20) with 137 choice questions between a real larger, later (LL) reward of $10 available after one of six delays (0, 7, 30, 90, 180, or 365 days, i.e., the temporal horizon ranged from 0 to 365 days) and a real smaller, sooner (SS) reward, which varied between $0.01 and $10.50, available immediately.[2] The questions were drawn randomly from this battery, without replacement, and presented to subjects sequentially. At the end of the experiment, one of a subject's choices was selected randomly for payment.

Mitchell used each subject's choices to determine an indifference point between the LL reward (i.e., $10) available after a particular delay (e.g., 7 days) and an SS reward available immediately. For example, if a subject chose $10 in 7 days over $6.50 immediately but then chose $7 immediately over $10 in 7 days, the subject was assigned an indifference point of $6.75. Taking the average of these two values is arbitrary and doing so throws away information about the uncertainty of this estimate; all that one can infer from this pattern of choices is that a subject's indifference point lies in the open interval ($6.50, $7). Interval data of this form is analysed appropriately using interval regression methods but Mitchell (1999) used the *estimated* indifference points as *data* to construct Mann-Whitney tests of differences in the indifference points of heavy smokers and never-smokers. Mitchell (1999) found that current smokers' indifference points were significantly lower than never-

---

[1] The smokers in Mitchell's (1999) study stated that they smoked at least 15 cigarettes per day and provided a breath sample to verify their smoking status.
[2] One would expect people to choose $10.50 now over $10 available after a delay so incorporating this question may be a test of subject comprehension; Mitchell (1999) provided no justification for the question's inclusion.

smokers' indifference points for the 7 day ($p < 0.05$), 30 day ($p < 0.01$), and 90 day ($p < 0.06$) delays.

In addition, Mitchell (1999) fitted Mazur's (1984) hyperbolic (H) discounting function to the indifference points for each subject and then compared the estimated discounting parameters of current smokers and never-smokers using a Mann-Whitney test; she found that current smokers discounted significantly more than never-smokers ($p < 0.06$). Using the point estimate of a discounting parameter as a datum ignores the uncertainty of this estimate and, thus, should not be used for inferential purposes.

Bickel, Odum and Madden (1999) (BOM) presented 23 heavy, current smokers, 22 never-smokers, and 21 ex-smokers ($N_S = 23$, $N_{NS} = 22$, and $N_{ES} = 21$)[3] with 189 choice questions between a hypothetical LL reward of $1000 available after one of seven delays (the temporal horizon ranged from 7 days to 25 years) and a hypothetical SS reward available immediately. For each delay, the SS rewards were presented sequentially in descending and then ascending order and subjects were asked to indicate their preference between each SS reward and the LL reward of $1000. A simple average of the last SS reward chosen in descending order and the first SS reward chosen in ascending order was used to define a person's indifference point for the $1000 LL reward at a particular delay; this method was used to derive 7 indifference points, for the seven delays in the task, for each subject even though taking the average of these two values is arbitrary and doing so throws away information about the uncertainty of the estimate.

BOM then fitted an exponential (E) function and Mazur's (1984) H function, using non-linear least squares (NLLS) estimation, to each subject's 7 derived indifference points. For each subject, BOM then compared[4] the fit of these two functions using the coefficient of determination, R^2, and found that for most subjects, H provided a better fit than E. They then used each subject's *estimated* R^2 value for the E and H functions as *data* to construct tests of whether the E or H functions provided a better fit to the

[3] The smokers in BOM (1999) reported smoking at least 20 cigarettes per day for 5 years and had a Fagerström Test for Nicotine Dependence score of at least 6. Never-smokers reported never smoking and ex-smokers reported abstinence for at least one year following 5 years of smoking at least 20 cigarettes a day.

[4] This was a simple comparison of the point estimates of R^2 for the E and H functions and not a formal statistical test.

discounting data across all subjects in the different smoking groups. Using the point estimate of a statistic (i.e., the value of R^2) as a datum ignores the uncertainty of this estimate and, thus, does not produce a valid test of one function's ability to better explain discounting data. Nevertheless, BOM state that the H function provided a better fit than the E function among current smokers ($p < 0.01$), never-smokers ($p < 0.01$), and ex-smokers ($p < 0.01$).

Finally, BOM compared the point *estimates* of the H discounting function across the current smoker, never-smoker, and ex-smoker groups by estimating an analysis of variance (ANOVA) model which included a smoking status covariate; they found a significant overall effect of smoking status but estimates should not be used as data for inferential purposes. In addition, planned Mann-Whitney pairwise comparisons of the H discounting function *estimates* showed that current smokers discounted significantly more than never-smokers ($p < 0.01$), and ex-smokers ($p < 0.01$); there was no significant difference between never-smokers and ex-smokers.

Thus, the first two studies analysing the relationship between smoking and discounting behaviour suggested that current smokers discount more heavily than never-smokers. In addition, it appeared that this result was robust to different subject pools, real as opposed to hypothetical rewards, different LL reward magnitudes ($10 versus $1000), and different elicitation mechanisms (random versus ordered choice). Although not confirmed by both studies, there was some evidence that the H function provided a better fit to the discounting data than the E function.

Reynolds, Karraker, Horn and Richards (2003) (RKHR), in a study with adolescent smokers ($N_S = 19$), adolescent never-smokers ($N_{NS} = 19$), and adolescent "triers" ($N_T = 17$)[5], provided the first null result in this literature. They used the titration procedure of Richards, Zhang, Mitchell and de Wit (1999) (RZMW), dubbed "Titration (random) – Richards et al. (1999)" in Table 1 in the main text, to derive indifference points for real $10 LL rewards available at different points in time (the temporal

[5] "Triers" had smoked cigarettes for the first time in the 6 months prior to the study and they smoked an average of 3.76 cigarettes in total over this time span. Smokers, by contrast, had smoked every week for at least 6 months prior to the study and they smoked 46.42 cigarettes, on average, per week.

horizon ranged from 1 to 365 days). This titration procedure has been used extensively in the smoking and discounting literature and deserves further comment.

A titration procedure uses a subject's choices to determine the next set of choices that the subject faces. As a simple example, if someone chooses $10 after 7 days over $5 now, the titration algorithm assumes that $10 after 7 days will be chosen over all amounts of money less than $5 available now (e.g., $4 or $3 available now). Consequently, the titration algorithm will narrow the search for a subject's indifference point for that delay period (i.e., 7 days in our example) to the open interval ($5, $10). Some titration procedures take the average of the two values defining that interval to determine the next SS reward presented to the subject, $7.50 in this case. If the subject chooses $7.50 now over $10 in 7 days, then the range for indifference points is narrowed to ($5, $7.50). If, by contrast, the subject chooses $10 in 7 days over $7.50 now, then the range for indifference points is narrowed to ($7.50, $10). By continually splitting the difference of an interval, the titration algorithm converges to a subject's indifference point for a particular delay.[6,7]

An issue with this titration procedure is that if a subject makes a mistake (e.g., chooses $10 after 7 days when he meant to choose $5 now), it becomes impossible to recover the subject's "true" indifference point because the algorithm uses that mistake to refine the subsequent set of choices presented to the subject. The algorithm developed by RZMW is more sophisticated and uses two top and two bottom limits, rather than one top limit (e.g., $10 in our previous example) and one bottom limit (e.g., $5 in our previous example), to alleviate this issue. By employing multiple top and bottom limits, the algorithm of RZMW can recover a subject's indifference point even after a mistake.

Another issue with a simple titration algorithm which splits the difference of an interval is that the adjusting nature of the algorithm is evident to the subject and he can deduce that his future choices depend on his current choices. This raises an

[6] Clearly the algorithm must terminate at some point, lest it continue indefinitely. In studies with $10 LL rewards, as in RKHR, the algorithm stopped when the difference between the rewards in the interval had declined to $0.50.
[7] As the interval within which a person's indifference point lies gets smaller and smaller, it is questionable whether that person is willing or able to make increasingly fine-grained choices.

obvious incentive-compatibility problem because the subject can "game" the algorithm so as to be presented with higher SS rewards on subsequent decisions. The algorithm of RZMW attempts to mitigate this problem by randomly drawing SS amounts from within an interval, rather than simply splitting the difference, and by randomly selecting LL reward delays, rather than determining the indifference point for one delay before moving on to the next.

RKHR used this algorithm to investigate the discounting behaviour of adolescent smokers, never-smokers, and "triers." Subjects also completed a probability discounting task and they were paid for one choice across both tasks; this payment scheme is referred to as 1-out-of-2-tasks in Table 1. The H discounting function was estimated for each subject, using NLLS, and the estimated discounting parameters were log transformed to normalise their distribution.[8] These transformed discounting parameters were used as data and fed into an ANOVA model so as to compare the discounting behaviour of the 3 smoking groups: there were no significant differences between smokers, never-smokers and "triers." As discussed in the main text, this near-universal two-step approach to data analysis is not valid statistically because point estimates are used as data in subsequent statistical models.

Table 1 collates the results from the other studies and, on inspection, a number of interesting patterns emerge. The vast majority (i.e., 25) of the studies investigating smoking and discounting behaviour were conducted in the US, 2 took place in Japan, 1 in Denmark, 1 in Canada, 1 in Germany, and 1 recruited subjects internationally over the internet. An important feature of these studies is that they have relatively diverse subject pools (i.e., they do not typically rely on convenient student samples but rather recruit from the community at large) which thereby bolsters the external validity of the results.

However, most of the studies have small sample sizes: 19 of the 31 studies recruited less than 100 people and 15 of these studies had samples of less than 70 people.

[8] RKHR used the common (i.e., base 10) logarithm to transform their data. A number of studies (e.g., Reynolds, Richards, Horn and Karraker (2004), Reynolds (2004), Heyman and Gibb (2006), Johnson, Bickel and Baker (2007) also adopt the common logarithmic transformation while others (e.g., Baker, Johnson and Bickel (2003), Bickel, Yi, Kowal and Gatchalian (2008), Jones, Landes, Yi and Bickel (2009), Sheffer et al. (2013)) use the natural logarithm to transform discounting parameters.

Fortunately, since 2008, the trend has been towards larger and larger samples (e.g., Sweitzer et al. (2008) recruited 710 subjects, Audrain-McGovern et al. (2009) recruited 909 subjects, Hofmeyr et al. (2017) recruited 1205 subjects, Kang and Ikeda (2014) used a sample of 3450 people, and Stillwell and Tunney (2012) recruited 9038 individuals).

With regard to elicitation mechanisms, 17 studies used choice procedures, 13 used titration, and 1 employed both methods (see Balevich, Wein and Flory (2013)). A perennial issue in the interpretation of experimental results is whether real or hypothetical rewards were used in a study. If a study uses hypothetical rewards, all it really elicits is the choices a person thinks he would make when faced with those contingencies, or the choices he thinks the experimenter wants him to make. If real rewards are used, by contrast, a subject's choices ultimately determine the payment he receives and this – coupled with a task that is easily understood, a transparent payment scheme, salient rewards, and an incentive-compatible experimental design – promotes truthful revelation of preferences. Thus, one should give far more credence to studies using real as opposed to hypothetical rewards because in the former instance one can analyse what people actually did rather than what they think they would do or what they want the experimenter to think they would do.

Of the studies in Table 1, 5 only used real rewards (Mitchell (1999), Reynolds et al. (2007), Melanko et al. (2009), Reynolds and Fields (2012), Kobiella et al. (2014)) whereas 4 used a combination of real and hypothetical rewards (Baker, Johnson and Bickel (2003), Heyman and Gibb (2006), Johnson, Bickel and Baker (2007), Mitchell and Wilson (2012)). Entirely hypothetical rewards were used in 15 studies (BOM, Ohmura, Takahashi and Kitamura (2005), Reynolds (2006), Bickel, Yi, Kowal and Gatchalian (2008), Sweitzer et al. (2008), Adams and Nettle (2009), Jones, Landes, Yi and Bickel (2009), Businelle, McVay, Kendzor and Copeland (2010), Bickel et al. (2012), Stillwell and Tunney (2012), Wing, Moss, Rabin and George (2012), Balevich, Wein and Flory (2013), Poltavski and Weatherly (2013), Sheffer et al. (2013), Kang and Ikeda (2014)), 1 study did not report whether real or hypothetical rewards were used (Audrain-McGovern et al. (2009)), and 6 studies used probabilistic payment schemes (RKHR, Reynolds (2004), Reynolds, Richards, Horn and Karraker (2004), Chabris et al. (2008), Harrison, Lau and Rutström (2010), Hofmeyr et al.

(2017)).[9] Thus, approximately half of the studies in Table 1 used entirely hypothetical rewards and this should be taken into account when drawing conclusions about the relationship between smoking and discounting behaviour.

The temporal horizon (i.e., the time delay between the SS and LL rewards) of the studies reported in Table 1 ranges from 6 hours to 25 years. Studies using real rewards or probabilistic payment schemes tend to employ far shorter temporal horizons than studies using hypothetical rewards; this makes sense because the credibility of payments in the distant future would be open to question. Of the studies using real rewards or probabilistic payment schemes, only 1 had a temporal horizon as long as 2 years (i.e., Harrison, Lau and Rutström (2010)) whereas 7 studies using hypothetical rewards had temporal horizons extending out to 25 years (BOM, Baker, Johnson and Bickel (2003), Johnson, Bickel and Baker (2007), Bickel, Yi, Kowal and Gatchalian (2008), Jones, Landes, Yi and Bickel (2009), Businelle, McVay, Kendzor and Copeland (2010), Sheffer et al. (2013)).

Time preferences are represented mathematically using a discounting function. There are a number of discounting functions which have been proposed but the majority of studies in Table 1 (i.e., 22 out of 31) adopted the assumption that people discount hyperbolically and, thus, only used Mazur's (1984) H function in their analyses.[10] There are 4 studies in Table 1 which directly compared the E and H discounting functions, using either R^2 or the residual sum of squares (RSS) to adjudicate between them, and all of the studies found that the H function better explains discounting data (BOM, Ohmura, Takahashi and Kitamura (2005), Bickel, Yi, Kowal and Gatchalian (2008), Stillwell and Tunney (2012)). There are 3 studies (Reynolds et al. (2007),

[9] Studies employing real rewards typically make use of the random lottery incentive mechanism (RLIM) to determine subject payment. RLIM randomly selects one of a subject's choices on a task and, in a study with real rewards, pays out this choice with certainty. A probabilistic payment scheme also makes use of RLIM but subjects are only given some chance of being paid for the randomly selected choice (i.e., subjects are not paid with certainty). In Chabris et al. (2008) subjects were given a 1-in-6 chance of being paid for one of their choices while in Harrison, Lau and Rutström (2010) subjects were given a 1-in-10 chance of being paid for one of their choices. By contrast, RKHR, Reynolds, Richards, Horn and Karraker (2004), Reynolds (2004) and Hofmeyr et al. (2017) paid subjects for 1 choice across two different tasks, implying that subjects had roughly a 50% chance of being paid for one of their choices on the discounting task.

[10] Businelle, McVay, Kendzor and Copeland (2010) and Poltavski and Weatherly (2013) assumed hyperbolic discounting but also used the area under the curve (AUC) method of Myerson, Green and Warusawitharana (2001) to draw inferences about the relationship between smoking and discounting. Mitchell and Wilson (2012) assumed hyperbolic discounting but also estimated a quasi-hyperbolic discounting function.

Melanko et al. (2009), Reynolds and Fields (2012)) which only used the "theoretically neutral" area under the curve (AUC) method of Myerson, Green and Warusawitharana (2001) to compare the discounting of smokers and non-smokers[11], and 2 studies (Harrison, Lau and Rutström (2010) and Hofmeyr et al. (2017)) estimated a statistical model that allows both E and H discounting functions to characterise the data.

The approach of Harrison, Lau and Rutström (2010) (HLR) and Hofmeyr et al. (2017) is based on the idea that some discounting choices may be better explained by an E function whereas others may be better explained by an H function and that the data should be used to determine the proportion of choices best explained by each model. Using this so-called "mixture model" approach, HLR found that approximately 25% - 40% of discounting choices were best characterised by the H function while Hofmeyr et al. (2017) found that 41% - 52% were best characterised by the H function. This suggests that researchers investigating the link between smoking and discounting behaviour may have relied too heavily on the H function because it does not explain all discounting choices all of the time.

Mixture models also address a deeper issue of bias in the estimation of discounting models. Suppose, for example, that at least 50% of the choices in a dataset are best characterised by the H function whereas the remaining fraction is best characterised by the E function. If one just estimates the H model on the whole dataset then one will reject the E model in favour of the H model because the estimate obtained from the H model will be halfway between the "true" H estimate and the estimate one would obtain from the E model. Thus, if one just estimates the H model then this biases against the E model. Mixture models remove this source of bias by allowing both discounting models to account for the data, and by estimating the proportion of the data which each model explains. We adopt the approach of HLR and Hofmeyr et al. (2017) and estimate mixture models of a number of different discounting specifications so as not to be wedded to any particular discounting framework and to

[11] The AUC method is "theoretically neutral" because it does not assume that discounting takes a particular form (e.g., E or H). Instead, when using the AUC method, one calculates the area under a subject's derived indifference points and normalises this to lie in the closed unit interval [0, 1]. Larger AUCs imply shallower discounting and, thus, the AUCs of smokers and non-smokers can be compared to determine whether the groups differ in their discounting behaviour.

determine the proportion of discounting choices that is explained by each specification.

HLR deserves further comment because it was the first study in this literature to use a front end delay (FED) to the SS reward and it is the only study which incorporates utility function curvature when estimating discounting models. Prior to the work of Coller and Williams (1999) it was common to make receipt of the SS reward immediate, as is the case in most of the studies in Table 1. An issue with an experimental design where the SS reward is immediate (i.e., a design with no FED) is that it may increase preference for the SS reward due to the additional transaction costs and uncertainty associated with receipt of the LL reward. A FED is used to hold these transactions costs constant across the two rewards. Following the work of HLR, Mitchell and Wilson (2012), Kang and Ikeda (2014) and Kobiella et al. (2014) used a FED for some of the choices they presented to subjects.

Time preferences are defined over time-dated utility flows, not flows of money. These are equivalent if a utility function is linear but Andersen, Harrison, Lau and Rutström (2008) showed that if a utility function is concave then the assumption of linearity will, for the same observed choices, bias the estimation of discounting parameters upwards. Thus, to draw accurate inferences about discounting behaviour it is important to incorporate utility function curvature in the estimation of discounting models. To our knowledge, HLR are the only researchers to incorporate the shape of the utility function when analysing the relationship between smoking and discounting behaviour. They used a risk preference task to determine the curvature of the utility function, under the assumption that expected utility theory characterised choices over risky prospects, which they then estimated jointly with the parameters of discounting models.

Incorporating utility function curvature had a marked effect on their results. Assuming linear utility, HLR found that both male and female smokers discount significantly more than their non-smoking counterparts. However, when the discounting models were estimated jointly with the curvature of the utility function, only male smokers discounted significantly more than male non-smokers; there was no statistically significant difference in the discounting behaviour of female smokers and non-smokers. The null result for women, under joint estimation, was driven by the fact that female smokers were significantly more risk averse (i.e., had significantly more curvature in their utility functions) than female non-smokers. By assuming linear utility, this difference in utility function curvature among women showed up as a difference in their discounting behaviour. Thus, to draw accurate inferences about smoking and discounting, it is crucial to jointly estimate utility function curvature and discounting parameters.

An important feature of the estimates presented in Table 1 is that 27 of the studies computed daily discount rates. This is common in the behavioural psychology literature but not in economics where annual discount rates are the norm. Of the remaining studies, 3 estimated annual rates, and 1 estimated weekly rates.

As discussed in the main text, the last column of Table 1 reports whether the researchers found a significant statistical relationship between smoking and discounting behaviour. Of the 37 reported findings in Table 1, 29 were positive and significant while the remaining 8 were null results. Thus, the bulk of findings in this literature – irrespective of whether real or hypothetical rewards, long or short temporal horizons, choice or titration elicitation mechanisms, small or large samples, and simple or complex statistical procedures were used – point to a positive relationship between smoking and discounting behaviour.

ADDITIONAL REFERENCES

RICHARDS, J. B., L. ZHANG, S. H. MITCHELL, AND H. DE WIT (1999): "Delay or Probability Discounting in a Model of Impulsive Behavior: Effect of Alcohol," *Journal of the Experimental Analysis of Behavior*, 71, 121-43.

APPENDIX B
[ONLINE WORKING PAPER]

As discussed in the main text, Table 2 provides a detailed summary of studies investigating the relationship between risk preferences and smoking behaviour. In this appendix we discuss Table 2 in more detail.

Mitchell (1999) conducted the first experimental study investigating the risk preferences of 20 relatively heavy[12], current smokers ($N_S = 20$) and 20 never-smokers ($N_{NS} = 20$). She presented subjects with 137 choice questions between a lottery which paid out \$10 with specific probabilities ($p = 0.1, 0.25, 0.5, 0.75, 0.9$, and 1) and \$0 with the complementary probability (i.e., the lottery (\$10, p; \$0, $1 - p$)), and a sure amount of money which varied between \$0.01 and \$10.50. The questions were drawn randomly from this battery, without replacement, and presented to subjects sequentially. At the end of the experiment, one of a subject's choices was selected randomly for payment.

Mitchell used each subject's choices to determine a certainty equivalent for the lottery (\$10, p; \$0, $1 - p$) at different values of p. For example, if a subject chose \$4 over the lottery (\$10, 0.5; \$0, 0.5) but then chose the lottery (\$10, 0.5; \$0, 0.5) over \$3.50, the subject was assigned a certainty equivalent of \$3.75. Taking the average of these two values is arbitrary and doing so throws away information about the uncertainty of this estimate; all that one can infer from this pattern of choices is that a subject's certainty equivalent lies in the open interval (\$3.50, \$4). Interval data of this form is analysed appropriately using interval regression methods but Mitchell used the *estimated* certainty equivalents (i.e., the point estimate \$3.75 in the example) as *data* to construct Mann-Whitney tests of differences in the certainty equivalents of heavy smokers and never-smokers; no significant differences between these groups were found.

[12] The smokers in Mitchell's (1999) study stated that they smoked at least 15 cigarettes per day and provided a breath sample to verify their smoking status.

In addition, Mitchell fitted the probability discounting (PD) probability weighting function (PWF)[13] to the certainty equivalents for each subject and then compared the *estimated* PWF parameters of smokers and never-smokers. She found evidence of risk aversion in both groups (i.e., $\gamma > 1$) but no significant differences in the risk preferences of smokers and never-smokers. Echoing the issue raised earlier, using the point estimate of any parameter as a datum ignores the uncertainty of this estimate and should not be used for inferential purposes.[14]

Reynolds, Karraker, Horn and Richards (2003) (RKHR) found that risk preferences differ according to smoking status but perhaps not in the way that would be expected, and with a statistical approach which is not valid. RKHR used the titration algorithm of Richards, Zhang, Mitchell and de Wit (1999), which was discussed in Appendix A, to elicit certainty equivalents for the lottery (\$10, p; \$0, $1 - p$) at different values of p among adolescent smokers ($N_S = 19$), adolescent never-smokers ($N_{NS} = 19$), and adolescent "triers" ($N_T = 17$).[15,16] Subjects also completed a delay discounting task and they were paid for one choice across both tasks; this payment scheme is referred to as 1-out-of-2-tasks in Table 2.

RKHR fitted the PD PWF to the *estimated* certainty equivalents, using non-linear least squares (NLLS) estimation, and then used the *estimated* PWF parameters as *data* in an ANOVA model so as to compare the three smoking status groups. For reasons outlined earlier, this statistical approach is not valid but RKHR report that they found evidence of risk aversion in all groups ($\gamma > 1$) and they found that "triers" were more

---

[13] The PD model is just Yaari's (1987) dual theory of choice under risk limited to a circumscribed class of lotteries and with a specific PWF: $\pi(p) = p / [p + \gamma(1 - p)]$; if $\gamma > 1$ this represents probability pessimism and risk aversion.

[14] The seventh column of Table 2 lists the statistical method that was adopted in each study and provides a binary summary judgement (i.e., valid or not valid), in parentheses, of whether the statistical approach was valid given the data obtained in the experiment. This binary summary judgement does not imply that the estimates which the researchers obtained were "wrong" but rather that the method used to derive the estimates was not appropriate for the data.

[15] "Triers" had smoked cigarettes for the first time in the 6 months prior to the study and they smoked an average of 3.76 cigarettes in total over this time span. Smokers, by contrast, had smoked every week for at least 6 months prior to the study and they smoked 46.42 cigarettes, on average, per week.

[16] As discussed in Appendix A, a titration algorithm is susceptible to being "gamed" by subjects because it narrows the search for the interval within which a person's certainty equivalent, for a particular value of p, lies by making the choices an experimental subject faces contingent on his prior choices. Thus, titration procedures lack incentive compatibility. Moreover, RKHR used the mid-point of the titration-derived interval as the person's certainty equivalent, even though any value within this interval is consistent with the data generating process (DGP). In other words, RKHR used an estimate as data, without taking into account the uncertainty of this estimate.

risk averse than smokers ($p < 0.05$) and never-smokers ($p < 0.05$); there were no significant differences between smokers and never-smokers. As discussed in the main text, this near-universal two-step approach to data analysis is not valid statistically because point estimates are used as data in subsequent statistical models.

Table 2 collates the results from the other studies. A clear majority of the studies (8 out of 11) were conducted in the US, with only one study a piece taking place in Japan, Denmark, and South Africa. An important feature of these studies is that only 3 use student subject pools while the rest recruit from the community at large; diverse samples help to bolster the external validity of the results so it is unfortunate that the statistical analyses in every study except Harrison, Lau and Rutström (2010) (HLR) hinder meaningful inferences.

The majority of studies on risk preferences and smoking behaviour have small sample sizes: the first 7 studies listed in Table 2 recruited less than 60 people. Fortunately, since 2008, 4 relatively large studies have taken place: Anderson and Mellor (2008) (AM) elicited risk preference data on 79 smokers and 898 non-smokers; HLR recruited 252 subjects; Szrek, Chao, Ramlagan and Peltzer (2012) (SCRP) used a sample of 351 individuals; and Poltavski and Weatherley (2013) recruited 182 people.

With regard to elicitation mechanisms, there is a roughly equal split between titration (6 out of 11 studies) and choice procedures. AM, HLR, and SCRP used an ordered choice elicitation mechanism, originally devised by Miller, Meyer and Lanzetta (1969) and refined by Holt and Laury (2002) (HL), which has been used extensively in the experimental economics literature on choice under risk, and thereby deserves further comment.[17] This elicitation procedure is referred to as a multiple price list (MPL).

In a MPL, subjects are given a table with 10 rows, and on each row they must choose between a "safe" and a "risky" lottery. In Table B:1, which is adapted from Table I in HL (p. 1645), Option A is the "safe" lottery because the range of the prizes is small (e.g., (\$2.00, p; \$1.60, $1 - p$)), and Option B is the "risky" lottery because the range of

[17] Harrison and Rutström (2008, p. 44-61) provide a detailed discussion of different risk preference elicitation mechanisms.

the prizes is large (e.g., (\$3.85, p; \$0.10, $1 - p$)). On row 1 of the table $p = 0.1$, and as you move down the table p increases by 0.1 on each row, implying that by row 10, $p = 1$. In the last 3 columns of the table we have included the expected value (EV) of Option A, the EV of Option B, and their difference, although this information is not usually provided to subjects.

In row 1 of Table B:1, the EV of Option A exceeds the EV of Option B but by row 5 the EV of Option B exceeds the EV of Option A. The logic behind this elicitation mechanism is that only a very risk loving subject would choose Option B (the "risky" lottery) on row 1, and only a very risk averse subject would choose Option A (the "safe" lottery) on row 9.[18] A risk neutral subject would switch from choosing Option A to Option B as the EV difference first changes sign (i.e., on row 5 of the table). Thus, if a subject switches to Option B before row 5 he is risk loving, if he switches to Option B on row 5 he is risk neutral, and if he switches to Option B after row 5 he is risk averse.

TABLE B:1: LOTTERY CHOICES IN THE HL RISK PREFERENCE EXPERIMENT

	Option A				**Option B**				EV^A	EV^B	Difference
Row	p	\$	p	\$	p	\$	p	\$	(\$)	(\$)	(\$)
1	0.1	2.00	0.9	1.60	0.1	3.85	0.9	0.10	1.64	0.48	1.17
2	0.2	2.00	0.8	1.60	0.2	3.85	0.8	0.10	1.68	0.85	0.83
3	0.3	2.00	0.7	1.60	0.3	3.85	0.7	0.10	1.72	1.23	0.50
4	0.4	2.00	0.6	1.60	0.4	3.85	0.6	0.10	1.76	1.60	0.16
5	0.5	2.00	0.5	1.60	0.5	3.85	0.5	0.10	1.80	1.98	-0.18
6	0.6	2.00	0.4	1.60	0.6	3.85	0.4	0.10	1.84	2.35	-0.51
7	0.7	2.00	0.3	1.60	0.7	3.85	0.3	0.10	1.88	2.73	-0.85
8	0.8	2.00	0.2	1.60	0.8	3.85	0.2	0.10	1.92	3.10	-1.18
9	0.9	2.00	0.1	1.60	0.9	3.85	0.1	0.10	1.96	3.48	-1.52
10	1	2.00	0	1.60	1	3.85	0	0.10	2.00	3.85	-1.85

Source: HL (p. 1645)

We can say even more about risk preferences by putting some parametric structure on the subjects' utility functions. Specifically, if we assume that subjects employ a power utility function $U(y) = y^r$, which displays constant relative risk aversion (CRRA), and

[18] As row 10 involves sure outcomes (i.e., $p = 1$) it is not relevant to risk preferences at all but is a good test of whether subjects understood the experiment because one would expect them to choose the larger sure outcome (e.g., \$3.85 from the example) over the smaller sure outcome (e.g., \$2.00 from the example). Harrison and Rutström (2009, p. 132) also advocate including a row 0 where the smaller outcome under each lottery (i.e., \$1.60 under Option A and \$0.10 under Option B) is received with certainty so as to "bracket" the MPL logic. In other words, if subjects can see that they should choose Option A on row 0 and Option B on row 10, then all they need to determine is the row on which they switch.

that they evaluate lotteries according to expected utility (EU) theory, then we can use a subject's pattern of choices on the MPL to define bounds on the risk preference parameter r.[19]

For example, suppose that a subject chose Option A on the first 5 rows of Table B:1 and then switched to Option B on row 6. To calculate the upper bound on r we solve the following equation:

$$0.5(\$2.00)^r + 0.5(\$1.60)^r = 0.5(\$3.85)^r + 0.5(\$0.10)^r \Leftrightarrow r \approx 0.85$$

This equation defines the value of r which makes a subject indifferent between the two lotteries on row 5. To calculate the lower bound on r we solve the following equation:

$$0.6(\$2.00)^r + 0.4(\$1.60)^r = 0.6(\$3.85)^r + 0.4(\$0.10)^r \Leftrightarrow r \approx 0.59$$

This equation defines the value of r which makes a subject indifferent between the two lotteries on row 6. Thus, if a subject chooses the Option A lottery on the first 5 rows and then switches to the Option B lottery on row 6, this pattern of choices implies a risk preference parameter r which lies in the open interval (0.59, 0.85). Interval data of this form is analysed appropriately using interval regression methods but AM and SCRP used the mid-point of these intervals as data to compare the risk preferences of smokers and non-smokers. As discussed previously, this approach throws away useful information on the uncertainty of the parameter estimates and violates the statistical assumptions of the second-stage models. Thus, the inferences drawn from these data are not valid statistically.

As discussed in the main text, a majority of the studies in Table 2 (8 out of 11) adopted the PD approach[20] to risk preferences, which defines risk aversion solely in terms of the shape of the PWF.[21] As subjective probability distortions drive risk

[19] Under EU theory the shape of a utility function determines attitudes toward risk. Using the power utility function above, $r > 1$ denotes risk loving behaviour, $r = 1$ denotes risk neutral behaviour, and $r < 1$ denotes risk aversion. If $r = 0$, $U(y) = \ln y$, and if $r < 0$, $U(y) = -y^r$, following Wakker (2008). AM, HLR and SCRP use a different parameterisation of the CRRA utility function: $U(y) = y^{(1-r)}/(1-r)$. Under this formulation, $r < 0$ denotes risk loving behaviour, $r = 0$ implies risk neutral behaviour, and $r > 0$ denotes risk aversion; if $r = 1$, $U(y) = \ln y$.

[20] The PD model is just Yaari's (1987) dual theory of choice under risk limited to a circumscribed class of lotteries and with a specific PWF: $\pi(p) = p \,/\, [p + \gamma(1 - p)]$; if $\gamma > 1$ this represents probability pessimism and risk aversion.

[21] Of these studies, 3 also employed the area under the curve (AUC) method of Myerson, Green and Warusawitharana (2001). When using the AUC method, one calculates the area under a subject's derived certainty equivalents and normalizes this to lie in the closed unit interval. Larger AUCs imply

preferences in the PD framework, it is surprising that 6 out of these 8 studies only used 5 probabilities in the elicitation task; the remaining two studies (Mitchell (1999) and Yi, Chase and Bickel (2007)) only used 6 and 7 probabilities, respectively. AM, HLR, and SCRP assumed that EU theory characterises choice under risk, so risk preferences are determined solely by the shape of the utility function. All of these studies used a MPL, which has 10 probabilities, and they assumed a CRRA utility function: specifically, $U(y) = y^{(1-r)} / (1-r)$. We allow risk preferences to be determined both by the shape of the utility function and the shape of the PWF so as to provide a bridge between prior studies in the literature. In addition, this allows us to explore whether smokers and non-smokers differ in the shape of their utility functions, the shape of their PWFs, or both.

In contrast to studies of smoking and discounting behaviour, there is a greater proportion of studies using real rewards or probabilistic payment schemes in the literature on smoking and risk preferences. Table 2 shows that 4 studies (Mitchell (1999), Reynolds et al. (2007), AM, SCRP) used only real rewards, whereas 3 studies used probabilistic payment schemes (RKHR, Reynolds, Richards, Horn and Karraker (2004), HLR).[22] The remaining 4 studies (Ohmura, Takahashi and Kitamura (2005), Reynolds (2006), Yi, Chase and Bickel (2007), Poltavski and Weatherly (2013)) used entirely hypothetical rewards. The use of real rewards or probabilistic payment schemes – coupled with a task that is easily understood, a transparent payment scheme, salient rewards, and an incentive-compatible experimental design – promotes the truthful revelation of preferences and, thus, far more credence should be given to the results from these studies than those which employ hypothetical rewards.

As discussed in the main text, the final column of Table 2 shows whether the studies found a significant statistical relationship between risk preferences and smoking

---

less risk aversion and, thus, the AUCs of smokers and non-smokers can be compared to determine whether the groups differ in their risk preferences. However, the AUC method does not allow one to derive valid statistical tests of the hypothesis of differences in risk preferences.

[22] Studies employing real rewards typically make use of the random lottery incentive mechanism (RLIM) to determine subject payment. RLIM randomly selects one of a subject's choices on a task and, in a study with real rewards, pays out this choice with certainty. A probabilistic payment scheme also makes use of RLIM but subjects are only given some chance of being paid for the randomly selected choice (i.e., subjects are not paid with certainty). In HLR subjects were given a 1-in-10 chance of being paid for one of their choices. By contrast, RKHR and Reynolds, Richards, Horn and Karraker (2004) paid subjects for 1 choice across 2 different tasks, implying that subjects had roughly a 50% chance of being paid for one of their choices on the risk preference task.

behaviour: the results are equivocal and, other than HLR, the statistical analyses are not valid. Null results were reported in 3 studies, positive results were reported in 5 studies, and negative results were reported in 3 studies. These conflicting results cut across different elicitation mechanisms, real and hypothetical rewards, different frameworks for choice under risk, and different methods of analysis. Thus, Table 2 shows that the relationship between risk preferences and smoking behaviour, or lack thereof, differs markedly across studies.

ADDITIONAL REFERENCES

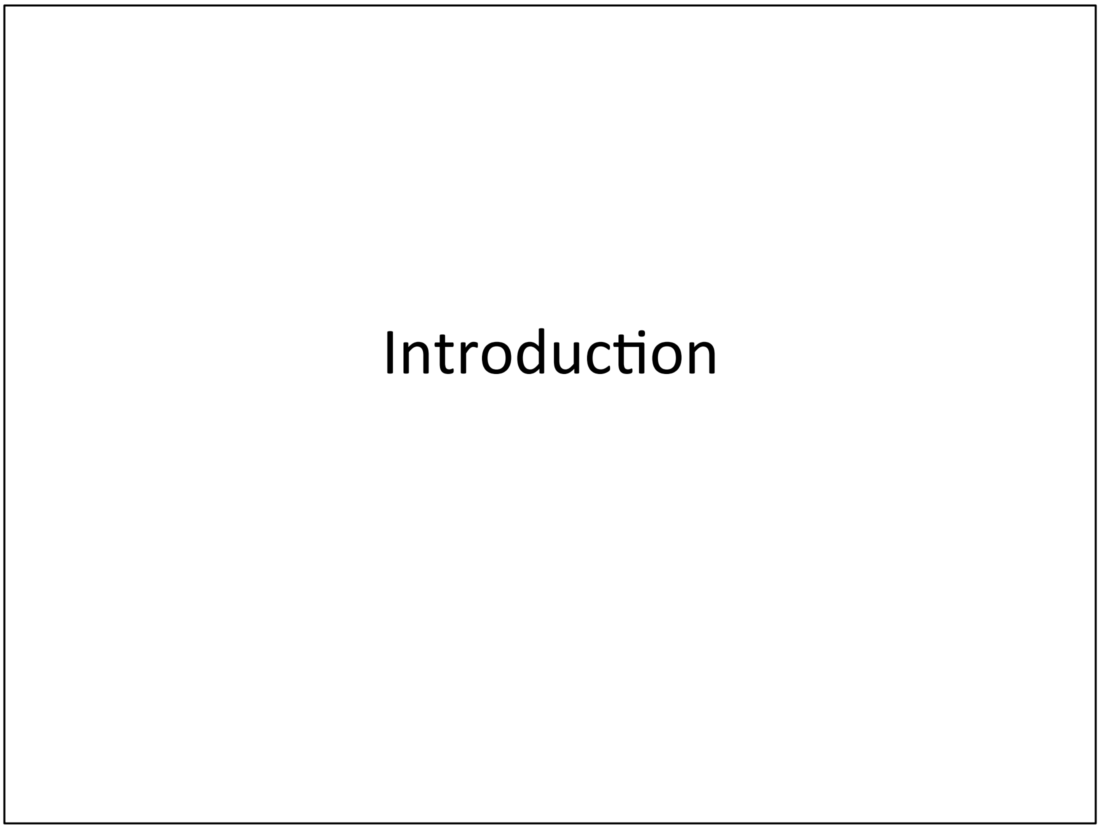
HARRISON, G. W., AND E. E. RUTSTRÖM (2008): "Risk Aversion in the Laboratory," in *Research in Experimental Economics: Volume 12. Risk Aversion in Experiments*, ed. by J. C. Cox, and G. W. Harrison. Bingley: Emerald, 41-196.

HOLT, C. A., AND S. K. LAURY (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644-1655.

MILLER, L., D. E. MEYER, AND J. T. LANZETTA (1969): "Choice among Equal Expected Value Alternatives: Sequential Effects of Winning Probability Level on Risk Preferences," *Journal of Experimental Psychology*, 79, 419-423.

APPENDIX C
[ONLINE WORKING PAPER]

The introductory presentation, risk preference task presentation, and time preference task presentation are included in this appendix. The introductory presentation explains the nature of the risk and time preference tasks and it includes a detailed discussion of the physical randomisation devices used in the experiments. The risk preference task presentation discusses the computer environment within which choices are made, the lotteries on offer and how to interpret them, and the payment scheme that is used to determine earnings. The time preference task presentation explains the computer environment within which choices are made, the calendar which shows subjects the dates at which smaller, sooner (SS) and larger, later (LL) rewards are available, and the payment scheme that is used to determine earnings. The presentations were designed to ensure that subjects understood how their choices ultimately led to the earnings they received so as to incentivise the truthful revelation of preferences.



Welcome

- Hello everyone and welcome to today's research session.
- Thank you for agreeing to take part in this study, your views and choices will be very informative and helpful.
- Before we get started I would like to explain how things are going to work.
- When I have finished this short explanation I will ask you to read and sign a consent form.
- Once that is done, we can begin with the tasks.

2 Tasks and a Questionnaire

- You will take part in 2 tasks and you will have the opportunity to earn money in each task.
- One of the tasks requires you to make choices between lotteries with varying prizes and chances of winning. You will make 40 of these choices.
- The other task asks you to choose between amounts of money available at different points in time. You will make 60 of these choices.
- Once you have completed the 2 tasks, you will need to fill out a short questionnaire.
- Once this is done, we will determine your earnings and you will be free to leave.

Earnings

- You will be paid R20 just for participating in today's session.
- At the end of the session, we will determine your earnings for the tasks.
- Some of this money will be paid to you at the end of the session and the rest of it will be paid to you in the future.
- This is why we need your bank details: to pay you via electronic transfer at a future date.
- To determine your earnings for the tasks, we will ask you to roll some dice.
- Let's go through a quick explanation of the dice you will roll.

10-sided dice

- At the end of today's session we will ask you to roll some dice into a plastic bowl which you can see below.
- Two of the dice that you will roll are 10-sided dice and these are used to select a number between 1 and 100.
- Every number between 1 and 100, and including 1 and 100, is equally likely to occur.
- An example of a dice roll is shown below.

10-sided dice

- Let's look at a close-up of the 10-sided dice.
- As you can see, one of the 10-sided dice has sides which increase in multiples of 10: 00, 10, 20, 30, 40, 50, 60, 70, 80 and 90.
- The other 10-sided die has sides which increase in multiples of 1: 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9.
- You will roll the two 10-sided dice together and add the numbers on the two dice to select a number between 1 and 100.
- In the example below, the number that was rolled is 86 (80 + 6).

10-sided dice

- To tell the difference between a 6 and a 9 there is a dot at the base of the number.
- This is why the number in the picture on the left below is a 6: there is a dot at the base of the 6.
- 9 looks different because there is a dot at the base of the 9.
- The picture on the right below shows you what a 9 looks like.

10-sided dice

- To roll a number between 1 and 9 you need to roll 00 and a single number between 1 and 9.
- As you can see in the picture on the left below, the number that was rolled is 5 (00 + 5).
- In the case where you roll 00 and 0, this will be treated as 100.
- As you can see in the picture on the right below, the number that was rolled is 100 (00 and 0).

Consent Form

- We have now finished the introductory explanation.
- To continue with today's session I need you to read and sign the consent form.
- This form explains your rights as a research participant and by signing it, you give your consent to participate in the study.
- If you have any questions please raise your hand and someone will come to answer them.
- You may read through the consent form now.

Task Instructions

Introduction

- In this task you will choose between lotteries with varying prizes and chances of winning.
- On each computer screen you will be presented with a pair of lotteries and you will need to choose one of them.
- There are 40 pairs of lotteries in this task.
- For each pair of lotteries, you should choose the lottery that you would prefer to play.
- You will actually get the chance to play <u>one</u> of the lotteries you choose, and you will be paid according to the outcome of this lottery.
- So you should think carefully about which lottery you prefer in each pair.

Computer Display

- All of the choices in this task will be made on a computer.
- This is what the computer display will look like:

- The display on your screen will be larger and easier to read.

Computer Display

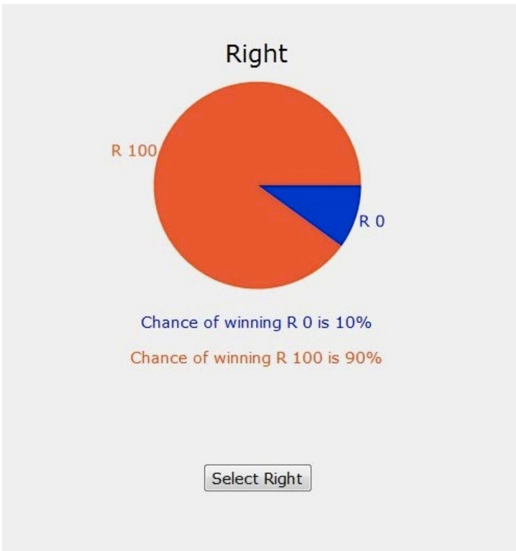
- On the computer screen there are two lotteries: a "Left" lottery and a "Right" lottery.
- Let's look at the Left lottery together.
- For the Left lottery there is a 20% chance of winning R0, a 60% chance of winning R100, and a 20% chance of winning R200.
- The coloured areas of the pie chart and the text below the pie chart represent these chances.

The Left Lottery

- As you can see, 20% of the pie chart is blue and this means you have a 20% chance of winning R0.
- This is what the blue text below the pie chart tells you: “Chance of winning R0 is 20%”.
- Similarly, 60% of the pie chart is red which means there is a 60% chance of winning R100.
- This is what the red text below the pie chart tells you: “Chance of winning R100 is 60%”.
- Finally, 20% of the pie chart is green which means there is a 20% chance of winning R200.
- This is what the green text below the pie chart tells you: “Chance of winning R200 is 20%”.

The Right Lottery

- If we look at the Right lottery we see that there is a 10% chance of winning R0 and a 90% chance of winning R100.
- 10% of the pie chart is blue and this means there is a 10% chance of winning R0.
- This is what the blue text below the pie chart tells you: “Chance of winning R0 is 10%”.
- 90% of the pie chart is red which means there is a 90% chance of winning R100.
- This is what the red text below the pie chart tells you: “Chance of winning R100 is 90%”.



Your Lottery Winnings

- The amount that you win from a lottery will be determined by the draw of a random number between 1 and 100.
- Each number between 1 and 100, and including 1 and 100, is equally likely to occur.
- You will draw this number yourself by rolling two 10-sided dice.
- One of the 10-sided dice has sides which increase in multiples of 10: 00, 10, 20, 30, 40, 50, 60, 70, 80 and 90.
- The other 10-sided die has sides which increase in multiples of 1: 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9.
- You will roll the two 10-sided dice together and add the numbers on the two dice to select a number between 1 and 100.
- For example, suppose the one 10-sided die lands on 70 and the other 10-sided die lands on 5.
- Then we will select number 75.
- We will work through an actual example of this later.

Choices

- Now, suppose that you prefer the Left lottery in the example below.
- To choose the Left lottery just click the button saying "Select Left".
- This is what the display will then look like if you choose the Left lottery.
- You can then click the button saying "Confirm" to move on to the next screen with a new pair of lotteries.
- If you would like to change your choice then just click "Cancel".

Choices

- Suppose instead that you prefer the Right lottery in the example below.
- To choose the Right lottery just click the button saying “Select Right”.
- This is what the display will then look like if you choose the Right lottery.
- You can then click the button saying “Confirm” to move on to the next screen with a new pair of lotteries.
- If you would like to change your choice then just click “Cancel”.

Total Number of Choices

- You will need to make 40 choices across 40 screens.
- On each screen there is a different lottery pair and you will need to choose either the Left lottery or the Right lottery.
- The Rand amounts under the lotteries change on each screen.
- In addition, the chances of winning the Rand amounts change for each lottery on each screen.
- So please pay careful attention when making each choice.
- At the end of the session today we will determine your earnings for this task in the following way.

Payment

- First, you will select one of the lottery pairs from this task by rolling a 4-sided die and then a 10-sided die.
- You will roll the 4-sided die to select 10 lottery pairs.
- If the die lands on 1, you will select lottery pairs 1-10; if the die lands on 2, you will select lottery pairs 11-20; if the die lands on 3, you will select lottery pairs 21-30; and if the die lands on 4, you will select lottery pairs 31-40.
- You will then roll the 10-sided die to select one lottery pair from this set of 10 pairs.
- For example, if the 4-sided die lands on 3, you will select lottery pairs 21-30.
- If you then roll a 7 on the 10-sided die, you will select lottery pair 27.
- Once you have selected the lottery pair, we will look at the choice that you made: the Left lottery or the Right lottery.
- We will then determine your winnings from this lottery by rolling two 10-sided dice, as explained earlier.
- Let's see what this means for the example we looked at earlier.

Payment

- Suppose that the lottery pair we looked at earlier gets selected for payment when you roll the 4-sided die and then the 10-sided die.
- And suppose that you chose the Left lottery on this screen.
- For the Left lottery there is a 20% chance of winning R0, a 60% chance of winning R100, and a 20% chance of winning R200.
- You will now roll two 10-sided dice to determine your winnings.
- As you can see on the screen, if you roll a number between 1 and 20, you will win R0. Thus, you have a 20% chance of winning R0.

Payment

- If you roll a number between 21 and 80 you will win R100. Thus, you have a 60% chance of winning R100.
- Finally, if you roll a number between 81 and 100 you will win R200. Thus, you have a 20% chance of winning R200.
- Suppose you roll the two 10-sided dice and one 10-sided die lands on 60 while the other 10-sided die lands on 7.
- Then we will select number 67.
- Because 67 is between 21 and 80, you will win R100.

Payment

- Thus, payment for this task is determined by three things:
  1. The lottery pair that is chosen to be played out using the 4-sided die and the 10-sided die.
  2. Your choice of the Left lottery or the Right lottery in each pair.
  3. The outcome of that lottery when you roll the two 10-sided dice.
- All winnings will be paid in cash at the end of today's session.

Choose the Option you Prefer

- The lottery you prefer in each pair is a matter of personal taste.
- The person next to you may have different tastes so their choices should not matter to you.
- Please work silently and make your choices by thinking carefully about each lottery.
- Since there is a chance that any one of your 40 choices could be selected for payment, you should approach each pair of lotteries as if it is the one that you will be paid for.
- If you have any questions please raise your hand and someone will come to answer them.
- You may begin the task.

Task Instructions

Introduction

- In this task you will choose between different amounts of money available at different times.
- You will need to make 60 choices in total.
- For each choice you will decide between a smaller amount of money which is available sooner and a larger amount of money which is available later.
- One of your 60 choices will be selected at random for payment and you will receive the amount of money that you chose on the appropriate date.

Computer Display

- All of these choices will be made on a computer.
- This is what the computer display will look like:

- The display on your screen will be larger and easier to read.

Computer Display

- At the top of the display is a calendar showing you today's date in purple (11 July 2012) and a future date in green (25 July 2012).
- Below the calendar are two columns: a purple column with amounts of money available today and a green column with amounts of money available in 14 days from today.
- You need to make 4 choices on this screen.
- Each choice appears on a different row.

Choices

- Let's look at the first row together.
- In the first row, you need to choose between receiving R200 today or R201.15 in 14 days from today.
- Note that R200 is the smaller of the two amounts but it is available today.
- R201.15 is the larger of the two amounts but it is only available after 14 days.
- Suppose that you prefer R200 today rather than R201.15 in 14 days.
- To choose R200 today just click the button saying "Select" under "R200 today".
- This is what the display will look like if you choose R200 today.

Choices

- Suppose instead that you prefer R201.15 in 14 days rather than R200 today.
- To choose R201.15 in 14 days just click the button saying "Select" under "R201.15 in 14 days".
- This is what the display will look like if you choose R201.15 in 14 days.
- Once you have made your choice on the first row you can move on to the other rows on the screen.
- Note that you need to make 4 choices on the screen before you can move on to the next set of 4 choices on a new screen.

Choices

- Once you have made all your choices on the screen, the display will look something like the example below.
- You can then click the button saying “Confirm” to move on to the next screen.
- If you would like to change your choices then click “Cancel”.

Total Number of Choices

- You will need to make 60 choices in total across 15 screens.
- The Rand amounts change on each row of each screen.
- In addition, the times for delivery of the Rand amounts change <u>across</u> the screens.
- For example, on the screen we just looked at, you had to choose between an amount of money available today and an amount of money available in 14 days.
- On a different screen, you may need to choose between an amount of money available in 14 days and another amount of money available in 21 days.
- So please pay careful attention when making your choices.
- At the end of the session today we will determine your earnings for this task in the following way.

Payment

- First, you will select one of the 15 screens from this task by rolling a 20-sided die.
- If the die lands on 1, you will select screen 1; if the die lands on 7, you will select screen 7; if the die lands on 12, you will select screen 12; and so on.
- If the die lands on 16, 17, 18, 19 or 20, you will roll the die again until it lands on a number between 1 and 15.
- Once you have selected a screen, you will roll a 4-sided die to select 1 of the 4 rows on the screen.
- If the die lands on 1, you will select row 1; if the die lands on 2, you will select row 2; and so on.
- Once you have selected the row, we will look at the choice you made on that row.
- You will then be paid for the choice that you made on that row on the date listed for that choice.
- Let's see what this means for the example we looked at earlier.

Payment

- Suppose that the screen we looked at earlier gets selected for payment when you roll the 20-sided die.
- You will then need to roll the 4-sided die to select a row for payment.
- Suppose that the die lands on 3, then you will select row 3.
- On row 3, you chose R209.70 in 14 days so you will be paid R209.70 in 14 days via electronic transfer.
- You will receive a payment confirmation when the transaction has taken place.

Choose the Option you Prefer

- Note that the option you prefer on each row is a matter of personal taste.
- The people next to you may have different tastes so their choices should not matter to you.
- Please work silently and make your choices by thinking carefully about each option.
- Since there is a chance that any of your 60 choices could be selected for payment, you should approach each choice as if it is the one that you will be paid for.
- If you have any questions please raise your hand and someone will come to answer them?
- You may begin the task.

APPENDIX D
[ONLINE WORKING PAPER]

The statistical method we employ is direct estimation by maximum likelihood of structural models of latent choice processes. The latent choice processes in question are captured by models of risk and time preferences. These models provide the structure necessary to estimate risk and time preferences using the observed choice data. One of the benefits of the maximum likelihood approach is that it uses all of the available information to estimate discounting and risk preference parameters and the precision of these estimates. We review the basic logic of the estimation strategy below, focussing on the canonical cases of EU theory and exponential (E) discounting. We then discuss the extension to other risk and time preference models.

Assume that utility of income is defined by a power utility function which displays constant relative risk aversion (CRRA):

$$U(y) = y^r, \tag{1}$$

where y is a lottery prize in the risk preference task and r is a parameter to be estimated. If $r = 0$, $U(y) = \ln y$, and if $r < 0$, $U(y) = -y^r$, following Wakker (2008). Under EU theory risk preferences are determined by the shape of the utility function, so with the power utility function parameterisation $r > 1$ yields a convex utility function and risk loving behaviour, $r = 1$ implies a linear utility function and risk neutrality, and $r < 1$ yields a concave utility function and risk aversion.

Let there be three possible outcomes in a lottery, just like the risk preference task used here. Under EU theory the probabilities for each outcome y_j, $p(y_j)$, are those that are used in the experimental task, so expected utility is the probability-weighted utility of each outcome in each lottery i:

$$\mathrm{EU}_i = \sum_{j=1,2,3} [\ p(y_j) \times U(y_j)\ ] \tag{2}$$

To determine the value of r, the EU for each lottery pair (i.e., the Left and Right lotteries in Figure 1) is calculated for a candidate estimate of r and an index of their differences is formed:

$$\nabla \mathrm{EU} = \mathrm{EU}_R - \mathrm{EU}_L \tag{3}$$

This is a latent index, based on latent preferences, which captures the difference in EU of the Right and Left lotteries presented to subjects. This index is then linked to the subjects' observed choices using the cumulative normal distribution function $\Phi(\nabla EU)$. This function takes any argument (∇EU) between $\pm\infty$ and smoothly transforms it into a number between 0 and 1. Thus, we have the so-called "probit" link function:

$$\Pr(\text{Choose lottery R}) = \Phi(\nabla EU) \tag{4}$$

The latent index in (3) is linked to subjects' observed choices by specifying that lottery R is chosen when $\Phi(\nabla EU) > ½$, which is what (4) implies.

The likelihood of the observed responses, conditional on the power utility and EU model being true, depends on the estimates of r given the statistical model above and the choices of subjects in the risk preference task. The conditional log-likelihood for the risk preference responses is:

$$\ln L_i^{RP}(r; z, X) = \sum_i[(\ln\Phi(\nabla EU)\times I(z_i=1)) + (\ln(1 - \Phi(\nabla EU))\times I(z_i=0))], \tag{5}$$

where $I(\cdot)$ is the indicator function, $z_i = 1(0)$ denotes the choice of the R(L) lottery in choice pair i, and X is a vector of individual characteristics capturing smoking status, age, gender, education etc.

One of the advantages of structural maximum likelihood estimation is that it is a straightforward extension to make the parameter of interest, the risk preference parameter r, a linear function of individual characteristics. In this case, one estimates $r = r_0 + r_\beta \times X$, where r_0 is a fixed parameter and r_β is a coefficient vector linked to the variable vector X of individual characteristics. If no individual characteristics are included in the model we estimate $r = r_0$, which is the risk preference parameter estimated at the level of the sample without taking into account observed, individual heterogeneity (i.e., assuming homogenous preferences). Every estimate of r includes a standard error which reflects our uncertainty as to the "true" value of r. This stands in sharp contrast to the bulk of studies in Table 2 which use risk preference point estimates as data in subsequent statistical models.

Another important extension to the simple model defined above is to allow for some behavioural errors on the part of subjects when they make choices between lotteries L and R. This error could be as simple as a "tremble," where, say, a subject wants to choose lottery R but mistakenly selects lottery L. We adopt the "contextual utility"

(CU) behavioural error specification of Wilcox (2011) to allow mistakes on the part of subjects from the perspective of the deterministic EU model and to draw robust inferences about the primitive "stochastically more risk averse than" relation.[23] The CU specification normalises the ∇EU index so that it falls within the closed unit interval [0, 1] and incorporates the behavioural error term originally due to Fechner (1966/1860). Thus, rather than adopt the simple ∇EU index in (3), we make use of the index:

$$\nabla EU = [(EU_R - EU_L) / \lambda] / \mu, \quad (6)$$

where λ is the normalising term and μ is the Fechner error term.

Different values of μ affect our ∇EU index. As $\mu \to 0$ this specification collapses to a deterministic choice model where the choice is strictly determined by the EU of the two lotteries. However, as $\mu \to \infty$, $\nabla EU \to 0$, and a subject's choice is essentially random, with an equal probability of selecting either lottery. When $\mu = 1$ we are back to specification (3), so the Fechner error term is a parameter which basically flattens the probit link function as its value increases. The new conditional log-likelihood is:

$$\ln L_i^{RP}(r, \mu; z, X) = \sum_i[(\ln\Phi(\nabla EU)\times I(z_i=1)) + (\ln(1 - \Phi(\nabla EU))\times I(z_i=0))] \quad (7)$$

The expression in (7) can be maximised using standard numerical methods to estimate the power function parameter r, which defines risk preferences under EU theory, and the Fechner error term μ, which determines the extent to which choices involve errors on the part of subjects.

It is a simple matter to incorporate other theories of choice under risk in this statistical framework. Quiggin (1982) developed the rank-dependent utility (RDU) model, which assumes that a decision maker transforms objective probabilities into subjective decision weights which are then used to evaluate lotteries. According to this theory, risk preferences are determined both by the shape of the utility function, like EU theory, and the shape of the PWF. Under RDU we replace (2) with:

$$RDU_i = \sum_{j=1,\ldots,n} [\, w(y_j) \times U(y_j) \,], \quad (8)$$

where

$$w_j = \pi(p_j + \ldots + p_n) - \pi(p_{j+1} + \ldots + p_n), \quad (9)$$

[23] The "stochastically more risk averse than" relation is the stochastic choice counterpart to the "more risk averse than" relation (see Pratt (1964)) which is defined for the deterministic EU model.

for $j = 1, \ldots, n$-1, and

$$w_j = \pi(p_j), \qquad (10)$$

for $j = n$. The subscript j represents outcomes ranked from worst to best, and $\pi(p)$ is a specific PWF.

A number of different PWFs have been used in the literature and Stott (2006) provides a useful review. Tversky and Kahneman (1992) (TK) popularised the following PWF:

$$\pi(p) = p^\gamma / [\, p^\gamma + (1\text{-}p)^\gamma \,]^{1/\gamma}, \qquad (11)$$

for $1 > p > 0$. This function permits linear, "inverse S-shaped" and "S-shaped" forms. Gonzalez and Wu (1999) review the empirical evidence on this function and find that $1 > \gamma > 0$ in most studies. This gives the function an inverse S-shape with overweighting of low probabilities up to a crossover point where $\pi(p) = p$, and then underweighting of moderate to high probabilities.[24] We estimate the TK PWF, amongst others, to see whether we replicate this inverse S-shaped result in this sample.

To estimate a RDU model, assuming power utility, the TK PWF, and the CU behavioural error specification, one forms the RDU index $\nabla\text{RDU} = [(\text{RDU}_\text{R} - \text{RDU}_\text{L})/\lambda]/\mu$ in the manner of (3) and then links this to the subjects' observed choices using the cumulative normal distribution function in the manner of (4). This defines the conditional log-likelihood for the model which is then used to estimate r, μ, and γ, where γ is the parameter defining the TK PWF. We estimate EU and RDU models to compare the risk preferences of smokers and non-smokers. In addition, we estimate the parameters of a variety of PWFs to ensure that the results are robust across different specifications.

Shifting to time preferences, under the E model, δ is the discounting parameter which equalises the *utility* of income received at time t with the *utility* of income received at time $t + \tau$:

$$[1 / (1 + \delta)^t]U(y_t) = [1 / (1 + \delta)^{t+\tau}]U(y_{t+\tau}), \qquad (12)$$

for some utility function $U(\cdot)$.

[24] However, Rieger and Wang (2006) and Ingersoll (2008) show that this function is not monotonic at very small values of γ.

Under the assumptions that EU characterises choices over risky prospects and that subjects employ the power utility function, we can add more structure to this indifference condition. Specifically, (12) becomes:

$$[1 / (1 + \delta)^t](y_t)^r = [1 / (1 + \delta)^{t+\tau}](y_{t+\tau})^r, \tag{13}$$

where the general form of the utility function $U(\cdot)$ in (12) has been replaced with the specific power utility function $U(y) = y^r$ in (13).

The left hand side of (13) represents the present value (PV) of the *utility* of the SS reward in the time preference task whereas the right hand side of (13) represents the present value of the *utility* of the LL reward. Thus,

$$\mathrm{PV}_{\mathrm{SS}} = [1 / (1 + \delta)^t](y_t)^r, \tag{14}$$

and

$$\mathrm{PV}_{\mathrm{LL}} = [1 / (1 + \delta)^{t+\tau}](y_{t+\tau})^r \tag{15}$$

To estimate the parameters of our time preference model, conditional on EU theory, power utility, and the E model, we form the latent index:

$$\nabla\mathrm{PV} = (\mathrm{PV}_{\mathrm{SS}} - \mathrm{PV}_{\mathrm{LL}}) / \nu, \tag{16}$$

where ν is a Fechner error term for the time preference task, just as μ was the behavioural error term for the risk preference task. We could force $\mu = \nu$ but there is little sense in doing so if we think that one task may be more cognitively challenging than the other, and hence more prone to subject error. To remain open to this possibility, we allow μ and ν to vary independently.[25]

The latent index (16) captures the difference in the present values of the utility of the SS and LL rewards. It is linked to subjects' observed choices using the cumulative normal distribution function $\Phi(\nabla\mathrm{PV})$. This defines our probit link function:

$$\Pr(\text{Choose SS reward}) = \Phi(\nabla\mathrm{PV}) \tag{17}$$

The latent index in (16) is linked to subjects' observed choices by specifying that the SS reward is chosen when $\Phi(\nabla\mathrm{PV}) > ½$, which is what (17) implies.

[25] Our prior is that the risk preference task, which incorporated up to three prizes in each lottery and a host of different probabilities, is more cognitively challenging than the time preference task, where subjects simply had to make choices between two rewards available at different points in time.

The likelihood of the observed time preference responses, conditional on the EU, power utility, and E models being true, depends on the estimates of r, δ, and ν, given the statistical model above. The conditional log-likelihood is:

$$\ln L_i^{TP}(r, \delta, \nu; z, X) = \sum_i[(\ln\Phi(\nabla PV)\times I(z_i=1)) + (\ln(1 - \Phi(\nabla PV))\times I(z_i=0))] \qquad (18)$$

The joint likelihood of the risk and time preference responses can then be formed as:

$$\ln L_i\,(r, \delta, \mu, \nu; z, X) = \ln L_i^{RP} + \ln L_i^{TP} \qquad (19)$$

This "joint estimation" approach, developed by Andersen, Harrison, Lau and Rutström (2008), uses subjects' choices in the risk preference task to pin down the parameters of the utility function, and subjects' choices in the time preference task to pin down the parameters of the E discounting model, conditional on the shape of the utility function. This approach ensures that we estimate time preferences defined over utility flows, and not flows of money.

It is straightforward to incorporate other discounting models in this statistical framework. In the case of Weibull discounting, for instance, (13) becomes:

$$[\exp(-\delta t^{(1/\beta)})](y_t)^r = [\exp(-\delta(t+\tau)^{(1/\beta)})](y_{t+\tau})^r \qquad (20)$$

(14) and (15) are adjusted appropriately to incorporate this new expression and then one forms the latent index in (16) and proceeds as before. Andersen, Harrison, Lau and Rutström (2014) review all of the major discounting models.

ADDITIONAL REFERENCES

GONZALEZ, R., AND G. WU (1999): "On the Shape of the Probability Weighting Function," *Cognitive Psychology*, 38, 129-166.

INGERSOLL, J. (2008): "Non-Monotonicity of the Tversky-Kahneman Probability-Weighting Function: A Cautionary Note," *European Financial Management*, 14, 385-390.

RIEGER, M. O., AND M. WANG (2006): "Cumulative Prospect Theory and the St. Petersburg Paradox," *Economic Theory*, 28, 655-679.

STOTT, H. P. (2006): "Cumulative Prospect Theory's Functional Menagerie," *Journal of Risk and Uncertainty*, 32, 101-130.

WAKKER, P. P. (2008): "Explaining the Characteristics of the Power (CRRA) Utility Family," *Health Economics*, 17, 1329-1344.

APPENDIX E
[ONLINE WORKING PAPER]

In this appendix we discuss the risk preference results under the assumptions that expected utility (EU) theory and rank-dependent utility (RDU) theory characterise choice under risk.

A. Risk Preferences: EU Theory

Table E:1 presents baseline estimates of an EU model employing a power utility function and the CU behavioural error specification. These results pool choices across all individuals, which means we are estimating the value of r_0 for the sample as a whole. In other words, we are initially assuming homogenous preferences. The results account for clustering at the individual level by adjusting the standard errors of the estimates to take into account the fact that each respondent made multiple choices across the 40 risk preference questions.

The estimate of r = 0.306 implies a relatively high level of risk aversion in the sample. The estimate of μ = 0.175 is positive and statistically significant, implying that subjects make behavioural errors in the risk preference task.

TABLE E:1: EXPECTED UTILITY THEORY ML ESTIMATES
HOMOGENOUS PREFERENCES

	Model
Power function parameter (r)	0.306***
	(0.028)
Error (μ)	0.175***
	(0.009)
N	7000
log-likelihood	-4198.932

Results account for clustering at the individual level
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

To analyse the link between risk preferences and smoking behaviour, one can make the parameter of interest r a linear function of smoking status. This captures the “total effect” of smoking status on risk preferences without controlling for any potential differences between smokers and non-smokers like age, education, and gender. The point estimate of the “smoker” variable in this model is -0.027 with a standard error of 0.048, which means there is not a statistically significant relationship between risk preferences and smoking status in this sample.[26]

TABLE E:2: EXPECTED UTILITY THEORY ML ESTIMATES
HETEROGENOUS PREFERENCES

	Model	
	Estimate	Std Error
Power function parameter (r)		
Age	0.005	0.012
White	0.045	0.05
Male	0.119**	0.047
Commerce faculty	0.081	0.055
Financial aid	-0.033	0.055
Risk task first	-0.031	0.046
Smoker	-0.036	0.058
Constant	0.123	0.223
Error (μ)		
Constant	0.173***	0.009
N	7000	
log-likelihood	-4180.528	

Results account for clustering at the individual level
* p<0.10, ** p<0.05, *** p<0.01

Table E:2 presents the results from a model that takes into account observed, individual heterogeneity by conditioning the power function parameter estimate on a set of covariates and task parameters. Specifically, the model includes the demographic variables from Table 3 in the main text and a variable specifying whether the risk preference task preceded the time preference task. This model captures the "marginal effect" of smoking status on risk preferences while controlling for other factors which may mediate this relationship.

Table E:2 shows that the only variable which is significantly and individually related to risk preferences in this sample is gender: men are less risk averse than women. Thus, estimates from the EU model with a power utility function and CU error specification point to no statistically significant differences in the risk preferences of smokers and non-smokers.[27]

[26] We also estimate a model which allows risk preferences to vary as a quadratic function of smoking intensity as measured by the number of cigarettes smoked per day: risk preferences are not significantly related to smoking intensity.

[27] To explore the possibility that the power utility function is too restrictive to accurately characterise choice under risk in this sample, we also estimate the expo-power (EP) utility function of Saha (1993), which admits increasing relative risk aversion (IRRA), decreasing relative risk aversion (DRRA), and CRRA. This EP utility function takes the following form: $U(y) = \theta - \exp(-\alpha y^r)$, where $\theta > 1$ and $\alpha r > 0$. In a model with the full set of covariates and task parameters, none of the coefficients nor the constant

B. Risk Preferences: RDU Theory

The EU results suggest that there are no significant differences in the risk preferences of smokers and non-smokers. However, this analysis, by assumption, ignored the role of probability weighting and it may be the case that smokers perceive probabilities differently to non-smokers. To explore this possibility, we estimate RDU models.

One of the key components of a RDU model is the specification of the PWF. The TK PWF was presented in Appendix D (Eq. 11). Two other commonly used PWFs are the power function and the Prelec (1998) function. The power PWF is just like the power utility function except that prizes are replaced with probabilities:

$$\pi(p) = p^{\gamma} \tag{1}$$

An important feature of the power PWF is that it is either concave, convex, or linear throughout its range. This means that interior probabilities are either viewed objectively (i.e., linear weighting), always overweighted, or always underweighted. Thus, the power PWF does not permit the inverse S-shaped or S-shaped forms of the TK PWF.

Prelec (1998) derived a two-parameter PWF which exhibits considerable flexibility. The functional form for this PWF is:

$$\pi(p) = \exp[-\eta(-\ln p)^{\varphi}], \tag{2}$$

which is defined for $1 > p > 0$, $\eta > 0$, and $\varphi > 0$.[28] This function allows independent specification of location and curvature in probability weighting. It also nests the power PWF when $\varphi = 1$, and nests a one-parameter function when $\eta = 1$, which is similar to the TK function and admits linear, inverse S-shaped, and S-shaped forms.

---

term for the expo parameter α are significantly different to zero. In addition, a test of the joint hypothesis that all of the covariates, including the constant term, are equal to zero for the expo parameter α cannot be rejected ($p = 0.833$). Harrison, Lau and Rutström (2007, p. 358) used this approach to determine whether CRRA held over the range of prizes used in their experiments. They too found that a test of the joint hypothesis that all of the covariates, and the constant term, are equal to zero, could not be rejected, which lead them to conclude that CRRA was an appropriate characterisation for their sample. Thus, the power utility function will be employed in subsequent analyses because it adequately characterises choice under risk in this sample.

[28] Prelec (1998, proposition 1, part C, p. 503) provides these parameter restrictions. Prelec (1998, proposition 1, part B, p. 503) constrains $1 > \varphi > 0$, but this constraint can be quite restrictive in practice because it restricts the PWF to be inverse-S shaped. When estimating the models we impose these constraints using nonlinear transformations of the parameters. To recover the core parameters we use the inverse of these nonlinear transformations, and then apply the "delta method" to derive standard errors and *p*-values for the estimates (see Oehlert (1992)).

Table E:3 presents baseline estimates of RDU models employing the power utility function, the CU behavioural error specification, and the three PWFs discussed above. In Model 1 the power PWF parameter $\gamma = 0.953$, implying slight overweighting of all probabilities. However, this estimate is not significantly different from 1 ($p = 0.301$) so we cannot rule out the hypothesis of a linear PWF where probabilities are viewed objectively.

TABLE E:3: RANK-DEPENDENT UTILITY THEORY ML ESTIMATES
HOMOGENOUS PREFERENCES

	Model 1	**Model 2**	**Model 3**
	Power	TK	Prelec
Power function parameter (r)	0.283***	0.351***	0.324***
	(0.024)	(0.032)	(0.026)
PWF parameter (γ/φ)	0.953***	0.868***	0.797***
	(0.045)	(0.022)	(0.025)
PWF parameter (η)			0.882***
			(0.033)
Error (μ)	0.176***	0.170***	0.169***
	(0.009)	(0.009)	(0.008)
N	7000	7000	7000
log-likelihood	-4197.975	-4177.421	-4151.295

Results account for clustering at the individual level
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

In Model 2 the TK PWF parameter $\gamma = 0.868$, which yields an inverse S-shaped function implying overweighting of low probabilities and underweighting of moderate to high probabilities. The estimate of γ is significantly less than 1 at any regular level of significance ($p < 0.001$).[29] In Model 3 we replicate the inverse S-shaped PWF that we find with the TK function as the estimates of $\varphi = 0.797$ and $\eta = 0.882$ are significantly less than 1 ($p < 0.001$ in both cases).

The estimates in Table E:3 show that probability weighting plays a role in the determination of risk attitudes in this sample. This will need to be taken into account when adopting the joint estimation approach to discounting behaviour because the extent of utility function curvature identified by the risk preference task propagates into estimates of discounting parameters. Thus, if one ignores probability weighting

[29] The presence of inverse S-shaped probability weighting explains why the estimate of γ is not significantly different to 1 in the model with the power PWF: in effect, the power PWF is “confused” because it has to be linear, concave, or convex throughout its range.

when it is present, this would lead to biased estimates of utility function curvature and, hence, biased estimates of discounting parameters. In effect, when probability weighting is present, one ought to apportion risk preferences into their concave utility and probability weighting components so that accurate inferences about discounting behaviour can be drawn.

Figure E:1: Estimated probability weighting functions and implied decision weights
Based on equi-probable reference lotteries

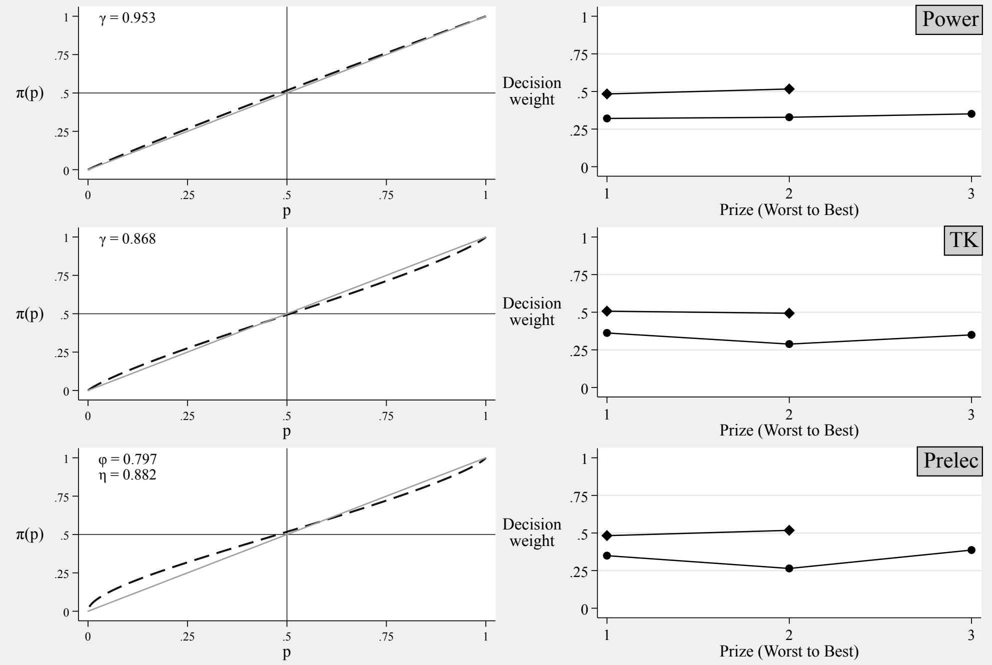


Figure E:1 plots the PWFs, and implied decision weights, for the estimates in Table E:3. The decision weights are graphed for equi-probable 2-outcome and 3-outcome reference lotteries. In the case of a 3-outcome equi-probable reference lottery, we show the decision weight applied to the worst outcome, the decision weight applied to the intermediate outcome, and the decision weight applied to the best outcome in the lottery.

To investigate the possibility that smokers perceive probabilities differently to non-smokers, even if their utility functions do not differ, we estimate the two models in Table E:3 which admit inverse S-shaped PWFs and allow the parameters to vary as a function of observable characteristics and task parameters. Results are presented in Table E:4.[30]

TABLE E:4: RANK-DEPENDENT UTILITY THEORY ML ESTIMATES
HETEROGENOUS PREFERENCES

	Model 1		Model 2	
	TK		Prelec	
	Estimate	Std Error	Estimate	Std Error
Power function parameter (r)				
Age	0.005	0.013	-0.004	0.011
White	0.038	0.060	0.029	0.051
Male	0.114**	0.055	0.062	0.049
Commerce faculty	0.113*	0.060	0.030	0.062
Financial aid	-0.057	0.065	-0.051	0.058
Risk task first	-0.057	0.055	-0.015	0.050
Smoker	-0.048	0.068	-0.005	0.055
Constant	0.179	0.246	0.366	0.230
PWF parameter (γ/φ)				
Age	-0.002	0.011	-0.003	0.006
White	0.021	0.054	0.001	0.047
Male	0.016	0.050	-0.009	0.044
Commerce faculty	-0.083	0.057	-0.084	0.120
Financial aid	0.061	0.059	0.034	0.056
Risk task first	0.055	0.051	0.054	0.080
Smoker	0.026	0.056	0.028	0.049
Constant	0.876***	0.228	0.871***	0.206
PWF parameter (η)				
Age			-0.027	0.046
White			-0.062	0.121
Male			-0.166	0.137
Commerce faculty			-0.216	0.184
Financial aid			-0.014	0.139
Risk task first			0.166	0.153
Smoker			0.146	0.177
Constant			1.425**	0.676
Error (μ)				
Constant	0.168***	0.008	0.166***	0.008
N	7000		7000	
log-likelihood	-4153.594		-4119.762	

Results account for clustering at the individual level
* p<0.10, ** p<0.05, *** p<0.01

[30] The experimental design of the risk preference task lends itself to common-ratio tests of EU theory. To complement the analyses in this section, we conduct a set of common-ratio tests for the lotteries represented in the MM triangles in Figure 2 to determine whether the choice patterns of smokers are more or less EU-consistent than non-smokers. We adopt the non-parametric Cochran Q test and find that both smokers and non-smokers violate EU theory in every MM triangle in Figure 2 ($p < 0.001$ in every test) except the MM triangle with a gradient of 3. In this latter MM triangle, we cannot reject the hypothesis that non-smokers satisfy EU theory ($p = 0.111$) but we can reject this hypothesis for smokers ($p = 0.027$). Thus, in only 1 of the 8 MM triangles of Figure 2 are non-smokers more EU-consistent than smokers. The bulk of the evidence, therefore, suggests little difference in the extent to which smokers and non-smokers violate EU theory; one reaches the same conclusion from the estimates in Table E:4.

In both of the models in Table E:4, smokers do not differ significantly from non-smokers in the shape of their utility functions (i.e., in the estimate of r) nor in the way they perceive probabilities (i.e., in the estimates of γ/φ and η). In addition, tests of the joint hypothesis that the coefficients for smokers across r, γ/φ, and η are equal to zero, cannot be rejected under either model ($p = 0.771$ for the TK model and $p = 0.823$ for the Prelec model).[31] Thus, at least in this sample, there are no significant differences in the risk preferences of smokers and non-smokers. This result is robust to different theories of choice under risk, different PWFs, and a utility function that admits varying relative risk aversion.

ADDITIONAL REFERENCES

HARRISON, G. W., M. I. LAU, AND E. E. RUTSTRÖM (2007): "Estimating Risk Attitudes in Denmark: A Field Experiment," *Scandinavian Journal of Economics*, 109, 341-369.

OEHLERT, G. W. (1992): "A Note on the Delta Method," *The American Statistician*, 46, 27-29.

SAHA, A. (1993): "Expo-Power Utility: A 'Flexible' Form for Absolute and Relative Risk Aversion," *American Journal of Agricultural Economics*, 75, 905-913.

[31] We also estimate a RDU model with the expo-power utility function, the Prelec PWF, and the full set of covariates from Table E:4. The smoker variable is not significantly different from zero for any of the parameters in the model. In addition, a test of the joint hypothesis that the coefficients for smokers across r, α, φ, and η are equal to zero, cannot be rejected ($p = 0.967$).

APPENDIX F
[ONLINE WORKING PAPER]

In this appendix we provide a more detailed discussion of the time preference models that were used in our empirical analyses. We also estimate the four discounting models used in the main text under the assumption that RDU theory characterises choice under risk so as to investigate the smoking-discounting relationship.

The discount factor for the E model is:

$$D^{E}(t) = 1 / (1 + \delta)^{t}, \quad (1)$$

for $t \geq 0$, and where the discount rate d is:

$$d^{E}(t) = \delta \quad (2)$$

Two important features of this model are that it is mathematically tractable (i.e., the geometric series $\sum_t D^{E}(t)$ converges in the limit), and the discount rate $d^{E}(t)$ is a constant over time which, when coupled with an additively-separable intertemporal utility function, implies time-consistent preferences.[32]

Phelps and Pollak (1968) developed the quasi-hyperbolic (QH) discounting function, in the context of a social planning problem, which has a discount factor:

$$D^{QH}(t) = 1 \qquad \text{if } t = 0 \quad (3a)$$

$$D^{QH}(t) = \beta / (1 + \delta)^{t} \qquad \text{if } t > 0 \quad (3b)$$

If $\beta = 1$ the QH specification collapses to the E model, whereas if $\beta < 1$ discounting is quasi-hyperbolic. We use β and δ to represent different parameters in each of the discounting models even though there is nothing which implies that they should be the same value across the different specifications; this choice was made for notational simplicity. When $\beta < 1$ discount rates decline over time in the QH model, which, under the assumption of an additively-separable intertemporal utility function, can yield time-inconsistent choices. Thus, the QH model can account for a "present-bias" or a "passion for the present" in discounting behaviour. Like the E discount factor, $\sum_t D^{QH}(t)$ converges in the limit

[32] Time consistency, or the lack thereof, is central to economic models of addiction. Time-inconsistent agents may fail to carry out plans they make for the future which provides a possible explanation for the behavioural puzzles listed earlier: 1) addicts expend resources to acquire their targets of addiction but then incur real costs to try to reduce or limit their consumption of these goods; and 2) the typical course of addiction is characterised by repeated unsuccessful attempts to quit prior to final abstention.

Mazur (1984, p. 427) developed the H discounting function to account for pigeons' preferences over fixed and variable schedules of reinforcement. The H specification has a discount factor:

$$D^{H}(t) = 1 / (1 + \delta t) \qquad (4)$$

This function has been used extensively in the psychology literature and in 28 of the 31 studies in Table 1. Unlike the E and QH discounting specifications, the harmonic series $\sum_t D^{H}(t)$ does not converge and the H model has not been widely used therefore in the theoretical economics literature. Mazur's (1984) H function forms part of a whole family of hyperbolic discounting models, but we use (4) due to its importance in the literature on time preferences and smoking.

The final discounting function which we estimate was originally proposed by Read (2001, equation 16, p. 25) and has been dubbed the "Weibull" (WB) discounting function by Jamison and Jamison (2011, p. 5) because it has an associated Weibull probability density function. The discount factor for the WB specification is:

$$D^{WB}(t) = \exp(-\delta t^{(1/\beta)}), \qquad (5)$$

for $\delta > 0$ and $\beta > 0$. When $\beta = 1$ (5) collapses to the E specification so the parameter β either "expands" or "contracts" time. When $\beta > 1$ it is as if time has contracted or is perceived to be "slowing down" by the individual, which yields declining discount rates and the potential for time inconsistency. By contrast, when $\beta < 1$ it is as if time has expanded or is "speeding up" as perceived by the individual. The individual is then assumed to behave "exponentially" with respect to these subjective perceptions of the time horizon.

Table F:1 presents results from the four discounting models employing the Fechner error term, assuming linear utility, and using years as the unit of measurement for the estimation of the parameters; this table is included for comparative purposes. Recall that if no individual characteristics are included in the model we estimate $\delta = \delta_0$ and $\beta = \beta_0$, which are the discounting parameters estimated at the level of the sample without taking into account observed, individual heterogeneity (i.e., assuming homogenous preferences). Recall that the results account for clustering at the individual level which adjusts the standard errors of the estimates to take into account the fact that each respondent made multiple choices across the 60 time preference questions.

TABLE F:1: DISCOUNTING FUNCTION ML ESTIMATES
LINEAR UTILITY AND HOMOGENOUS PREFERENCES

	Model 1	**Model 2**	**Model 3**	**Model 4**
	Exponential	Hyperbolic	Quasi-Hyperbolic	Weibull
Discounting parameter (δ)	3.234***	1.715***	2.833***	0.890***
	(0.287)	(0.096)	(0.271)	(0.066)
Discounting parameter (β)			0.962***	1.518***
			(0.013)	(0.107)
Error (ν)	24.272***	24.043***	23.669***	23.573***
	(1.774)	(1.742)	(1.626)	(1.591)
N	10500	10500	10500	10500
log-likelihood	-5419.508	-5335.484	-5352.777	-5233.649

Results account for clustering at the individual level
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Under the assumption of linear utility, estimated discount rates are huge and differ markedly across the different specifications. In the E model, the estimate of δ = 3.234 implies an annual discount rate in excess of 320%.

In Table F:2, by contrast, where we estimate the discounting models under the assumptions of RDU and the Prelec (1998) PWF, discount rates are far lower, and more similar. For example, the estimate of the E discount rate δ = 0.493 implies an annual discount rate of approximately 49%. Similar declines are evident across all of the discounting specifications which highlights the point, now familiar from Andersen, Harrison, Lau and Rutström (2008), that incorporating concavity of the utility function leads to substantial declines in inferred discount rates.

In the QH model, the estimate of β = 0.988, which captures a "present-bias" or a "passion for the present" in discounting behaviour, is statistically significantly less than 1 (p = 0.002), which provides evidence of quasi-hyperbolic discounting and declining discount rates. The same is true in the WB results: the estimate of β = 1.611, which "expands" or "contracts" time, is statistically significantly greater than 1 (p < 0.001) which leads us to infer that people perceive time as "slowing down," generating declining discount rates. Thus, both the QH and WB results suggest that discount rates decline over time, which, when coupled with an additively-separable intertemporal utility function, raises the spectre of time-inconsistent choices. However, the two discounting functions provide competing explanations for this result: a present-bias in the case of the QH model and subjective time perception in the case of the WB model.

TABLE F:2: DISCOUNTING FUNCTION ML ESTIMATES
RANK-DEPENDENT UTILITY AND HOMOGENOUS PREFERENCES

	Model 1	**Model 2**	**Model 3**	**Model 4**
	Exponential Prelec	Hyperbolic Prelec	Quasi-Hyperbolic Prelec	Weibull Prelec
Power function parameter (r)	0.277***	0.327***	0.260***	0.238***
	(0.028)	(0.027)	(0.029)	(0.030)
PWF parameter (φ)	0.797***	0.797***	0.796***	0.795***
	(0.025)	(0.025)	(0.025)	(0.026)
PWF parameter (η)	0.838***	0.884***	0.823***	0.804***
	(0.032)	(0.034)	(0.032)	(0.031)
Discounting parameter (δ)	0.493***	0.502***	0.415***	0.204***
	(0.062)	(0.050)	(0.057)	(0.028)
Discounting parameter (β)			0.988***	1.611***
			(0.004)	(0.115)
Risk error (μ)	0.178***	0.169***	0.181***	0.186***
	(0.009)	(0.008)	(0.010)	(0.010)
Time error (ν)	0.151***	0.231***	0.128***	0.104***
	(0.041)	(0.055)	(0.036)	(0.031)
N	17500	17500	17500	17500
log-likelihood	-9471.828	-9441.151	-9383.297	-9234.32

Results account for clustering at the individual level
Standard errors in parentheses
* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Tables F:3:A and F:3:B present the results from the four discounting functions where the parameters are allowed to vary as a function of smoking status, other observable characteristics, and task parameters; these models, therefore, capture the "marginal effect" of smoking status. Across all of the models, the estimate of δ for smokers is positive and statistically significant at the 1% level. By contrast, in the QH and WB models the estimate of β for smokers is not statistically significant.

Thus, we observe a positive relationship between smoking and discounting behaviour which holds across all of the time preference models and all of the model specifications we estimate. This result is also robust to the assumption that EU characterises choice under risk (see Appendix G). However, smokers do not differ from non-smokers with regard to present-bias in the QH model nor in terms of time perception in the WB model.

TABLE F:3:A: DISCOUNTING FUNCTION ML ESTIMATES
RANK-DEPENDENT UTILITY AND HETEROGENOUS PREFERENCES

	Model 1		**Model 2**	
	Exponential		Hyperbolic	
	Estimate	Std error	Estimate	Std error
Power function parameter (r)				
Age	-0.008	0.006	-0.009	0.006
White	-0.016	0.020	-0.020	0.022
Male	-0.019	0.017	-0.019	0.018
Commerce faculty	0.008	0.020	0.008	0.022
Financial aid	0.043**	0.019	0.046**	0.021
Risk task first	0.006	0.018	0.009	0.019
Smoker	0.060***	0.022	0.067***	0.024
Constant	0.443***	0.125	0.528***	0.131
PWF parameter (φ)				
Age	-0.003	0.013	-0.003	0.012
White	-0.019	0.062	-0.020	0.059
Male	-0.006	0.056	-0.003	0.053
Commerce faculty	-0.108	0.071	-0.108	0.068
Financial aid	0.014	0.059	0.005	0.056
Risk task first	0.090*	0.052	0.083*	0.050
Smoker	0.022	0.061	0.021	0.058
Constant	0.890***	0.275	0.882***	0.265
PWF parameter (η)				
Age	-0.022	0.019	-0.023	0.020
White	-0.125	0.087	-0.134	0.089
Male	-0.218***	0.081	-0.216***	0.082
Commerce faculty	-0.202**	0.102	-0.209**	0.104
Financial aid	0.061	0.086	0.059	0.087
Risk task first	0.130*	0.078	0.133*	0.080
Smoker	0.158*	0.096	0.168*	0.098
Constant	1.458***	0.389	1.550***	0.398
Discounting parameter (δ)				
Age	-0.003	0.016	-0.001	0.015
White	-0.101	0.074	-0.095	0.064
Male	0.132**	0.064	0.124**	0.052
Commerce faculty	0.032	0.078	0.018	0.068
Financial aid	0.121	0.075	0.095	0.063
Risk task first	0.024	0.066	0.031	0.058
FED: 1 week	0.059	0.071	0.058	0.062
FED: 2 weeks	-0.004	0.072	0.004	0.063
High Principal	-0.208***	0.036	-0.191***	0.026
Smoker	0.260***	0.070	0.220***	0.062
Constant	0.512	0.320	0.490*	0.289

Table continues on next page

TABLE F:3:A: DISCOUNTING FUNCTION ML ESTIMATES (CONTINUED)
RANK-DEPENDENT UTILITY AND HETEROGENOUS PREFERENCES

	Model 1		**Model 2**	
	Exponential		Hyperbolic	
	Estimate	Std error	Estimate	Std error
Risk error (μ)				
Constant	0.169***	0.008	0.159***	0.007
Time error (ν)				
Constant	0.196***	0.048	0.338***	0.070
N	17500		17500	
log-likelihood	-9076.945		-9024.286	

Results account for clustering at the individual level
* $p<0.10$, ** $p<0.05$, *** $p<0.01$

TABLE F:3:B: DISCOUNTING FUNCTION ML ESTIMATES
RANK-DEPENDENT UTILITY AND HETEROGENOUS PREFERENCES

	Model 3		**Model 4**	
	Quasi-Hyperbolic		Weibull	
	Estimate	Std error	Estimate	Std error
Power function parameter (r)				
Age	-0.007	0.005	-0.004	0.005
White	-0.019	0.019	-0.013	0.016
Male	-0.021	0.017	-0.025*	0.015
Commerce faculty	0.007	0.017	0.013	0.016
Financial aid	0.035*	0.019	0.025	0.017
Risk task first	-0.002	0.023	-0.001	0.015
Smoker	0.060***	0.022	0.049**	0.022
Constant	0.414***	0.111	0.339***	0.116
PWF parameter (φ)				
Age	-0.003	0.013	-0.004	0.013
White	-0.019	0.063	-0.018	0.064
Male	-0.006	0.057	-0.006	0.058
Commerce faculty	-0.109	0.072	-0.11	0.073
Financial aid	0.017	0.06	0.023	0.061
Risk task first	0.093*	0.053	0.094*	0.054
Smoker	0.022	0.062	0.024	0.063
Constant	0.892***	0.276	0.898***	0.279
PWF parameter (η)				
Age	-0.021	0.019	-0.018	0.02
White	-0.129	0.087	-0.12	0.087
Male	-0.220***	0.08	-0.225***	0.081
Commerce faculty	-0.203**	0.101	-0.194*	0.1
Financial aid	0.055	0.085	0.048	0.084
Risk task first	0.122	0.08	0.123	0.077
Smoker	0.157	0.097	0.144	0.098
Constant	1.429***	0.387	1.349***	0.403

Table continues on next page

TABLE F:3:B: DISCOUNTING FUNCTION ML ESTIMATES (CONTINUED)
RANK-DEPENDENT UTILITY AND HETEROGENOUS PREFERENCES

	Model 3		**Model 4**	
	Quasi-Hyperbolic		Weibull	
	Estimate	Std error	Estimate	Std error
Discounting parameter (δ)				
Age	-0.007	0.013	0.007	0.007
White	-0.088	0.071	-0.049	0.031
Male	0.142**	0.06	0.053***	0.02
Commerce faculty	0.018	0.075	-0.003	0.03
Financial aid	0.103	0.067	0.001	0.026
Risk task first	-0.044	0.065	-0.027	0.028
FED: 1 week	0.344***	0.079	0.139**	0.068
FED: 2 weeks	0.287***	0.066	0.204**	0.088
High Principal	-0.156***	0.03	-0.056***	0.015
Smoker	0.224***	0.073	0.082***	0.028
Constant	0.279	0.266	0.003	0.133
Discounting parameter (β)				
Age	-0.004	0.004	-0.007	0.079
White	-0.002	0.012	0.285	0.306
Male	0.011	0.009	-0.492	0.357
Commerce faculty	0.002	0.012	-0.032	0.238
Financial aid	-0.005	0.012	0.520*	0.284
Risk task first	-0.021*	0.011	0.929**	0.4
FED: 1 week	0.348	0.273	2.840*	1.45
FED: 2 weeks	0.167	0.217	4.155*	2.39
High Principal	0.006**	0.002	0.081	0.107
Smoker	-0.002	0.012	-0.382	0.701
Constant	1.040***	0.067	2.161	1.709
Risk error (μ)				
Constant	0.172***	0.008	0.175***	0.01
Time error (ν)				
Constant	0.195***	0.051	0.130***	0.047
N	17500		17500	
log-likelihood	-8826.297		-8522.212	

Results account for clustering at the individual level
* p<0.10, ** p<0.05, *** p<0.01

ADDITIONAL REFERENCES

JAMISON, D. T., AND J. JAMISON (2011): "Characterizing the Amount and Speed of Discounting Procedures," *Journal of Benefit-Cost Analysis*, 2, 1-53.

READ, D. (2001): "Is Time-Discounting Hyperbolic or Sub-Additive?," *Journal of Risk and Uncertainty*, 23, 5-32.

APPENDIX G
[ONLINE WORKING PAPER]

In this appendix, we replicate the results presented in the main text and Appendix F under the assumption that EU theory characterises choice under risk. In Table G:1 we estimate the discounting models jointly with the curvature of the utility function and replicate the result from Table F:2: discount rates are far lower, and more similar, when compared to the model assuming linear utility (Table F:1). In addition, in the QH model in Table G:1, the estimate of $\beta = 0.987$ is statistically significantly less than 1 ($p = 0.003$), which provides evidence of quasi-hyperbolic discounting and declining discount rates. The same is true in the WB results: the estimate of $\beta = 1.608$ is statistically significantly greater than 1 ($p < 0.001$) which leads us to infer that people perceive time as "slowing down" and this generates declining discount rates.

TABLE G:1: DISCOUNTING FUNCTION ML ESTIMATES
CONCAVE UTILITY AND HOMOGENOUS PREFERENCES

	Model 1	**Model 2**	**Model 3**	**Model 4**
	Exponential	Hyperbolic	Quasi-Hyperbolic	Weibull
Power function parameter (r)	0.283***	0.309***	0.273***	0.260***
	(0.032)	(0.028)	(0.034)	(0.037)
Discounting parameter (δ)	0.507***	0.472***	0.441***	0.223***
	(0.070)	(0.049)	(0.065)	(0.032)
Discounting parameter (β)			0.987***	1.608***
			(0.004)	(0.114)
Risk error (μ)	0.183***	0.174***	0.187***	0.192***
	(0.011)	(0.009)	(0.012)	(0.014)
Time error (ν)	0.159***	0.198***	0.145***	0.128***
	(0.047)	(0.050)	(0.045)	(0.044)
N	17500	17500	17500	17500
log-likelihood	-9519.026	-9488.92	-9430.8	-9282.495

Results account for clustering at the individual level
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Tables G:2:A and G:2:B analyse the smoking-discounting relationship by making the parameters of interest a linear function of observable characteristics and task parameters. In the E and H models in Table G:2:A, there is a positive and significant relationship between smoking and discounting behaviour: smokers discount the future more heavily than non-smokers.

Similarly, in the QH and WB models in Table G:2:B, the estimate of δ for smokers is positive and statistically significant. However, smoking status is not significantly related to the extent of present-bias in the QH model nor in the way people perceive time in the WB model (i.e., in the estimates of β). Thus, the results in tables G:2:A and G:2:B replicate those reported in the main text; the comparable tables are presented in Appendix F.

TABLE G:2:A: DISCOUNTING FUNCTION ML ESTIMATES
CONCAVE UTILITY AND HETEROGENOUS PREFERENCES

	Model 1 Exponential		**Model 2** Hyperbolic	
	Estimate	Std error	Estimate	Std error
Power function parameter (r)				
Age	-0.006	0.005	-0.007	0.005
White	-0.012	0.018	-0.014	0.019
Male	-0.012	0.016	-0.010	0.017
Commerce faculty	0.011	0.018	0.011	0.020
Financial aid	0.039**	0.018	0.040**	0.019
Risk task first	0.004	0.016	0.006	0.017
Smoker	0.052***	0.020	0.056***	0.021
Constant	0.382***	0.110	0.419***	0.111
Discounting parameter (δ)				
Age	-0.004	0.014	-0.003	0.011
White	-0.091	0.066	-0.081	0.052
Male	0.114**	0.055	0.098**	0.042
Commerce faculty	0.027	0.067	0.014	0.055
Financial aid	0.111*	0.067	0.085	0.052
Risk task first	0.022	0.058	0.025	0.047
FED: 1 week	0.053	0.063	0.048	0.051
FED: 2 weeks	-0.004	0.064	0.002	0.052
High Principal	-0.178***	0.033	-0.149***	0.021
Smoker	0.232***	0.062	0.187***	0.050
Constant	0.485*	0.276	0.444**	0.224
Risk error (μ)				
Constant	0.181***	0.011	0.170***	0.008
Time error (ν)				
Constant	0.155***	0.042	0.201***	0.044
N	17500		17500	
log-likelihood	-9163.252		-9117.061	

Results account for clustering at the individual level
* *p*<0.10, ** *p*<0.05, *** *p*<0.01

TABLE G:2:B: DISCOUNTING FUNCTION ML ESTIMATES
CONCAVE UTILITY AND HETEROGENOUS PREFERENCES

	Model 3		**Model 4**	
	Quasi-Hyperbolic		Weibull	
	Estimate	Std error	Estimate	Std error
Power function parameter (r)				
Age	-0.006	0.005	-0.003	0.005
White	-0.015	0.018	-0.010	0.015
Male	-0.013	0.016	-0.019	0.014
Commerce faculty	0.010	0.016	0.016	0.016
Financial aid	0.031*	0.018	0.023	0.016
Risk task first	-0.004	0.021	-0.002	0.015
Smoker	0.053***	0.020	0.044**	0.020
Constant	0.368***	0.101	0.307***	0.101
Discounting parameter (δ)				
Age	-0.007	0.011	0.007	0.007
White	-0.082	0.065	-0.047	0.029
Male	0.132**	0.056	0.051***	0.020
Commerce faculty	0.018	0.069	-0.002	0.029
Financial aid	0.094	0.063	0.000	0.025
Risk task first	-0.041	0.060	-0.026	0.027
FED: 1 week	0.313***	0.075	0.133**	0.067
FED: 2 weeks	0.265***	0.064	0.203**	0.092
High Principal	-0.139***	0.029	-0.054***	0.013
Smoker	0.206***	0.067	0.080***	0.026
Constant	0.260	0.242	0.005	0.129
Discount parameter (β)				
Age	-0.003	0.003	-0.009	0.079
White	-0.002	0.011	0.288	0.306
Male	0.011	0.009	-0.499	0.360
Commerce faculty	0.002	0.011	-0.026	0.239
Financial aid	-0.005	0.011	0.526*	0.281
Risk task first	-0.020*	0.010	0.922**	0.397
FED: 1 week	0.360	0.279	2.839*	1.452
FED: 2 weeks	0.140	0.213	4.379*	2.455
High Principal	0.005**	0.002	0.076	0.106
Smoker	-0.002	0.011	-0.382	0.706
Constant	1.035***	0.062	2.200	1.712
Risk error (μ)				
Constant	0.184***	0.011	0.187***	0.012
Time error (ν)				
Constant	0.163***	0.046	0.119***	0.039
N	17500		17500	
log-likelihood	-8912.317		-8606.675	

Results account for clustering at the individual level
* p<0.10, ** p<0.05, *** p<0.01

Table G:3 maps out the response surface for estimates of δ in the four time preference models evaluated at different values of number of cigarettes smoked per day and assuming EU theory characterises choice under risk. Analogous to Table 5 in the main text, at low values of number of cigarettes, the conditional marginal effect of additional cigarettes is positive. By 15 cigarettes, though, the conditional marginal effect of additional cigarettes is negative. Table G:3 therefore replicates the results from Table 5 in the main text and highlights the nonlinear effect of smoking intensity on discounting behaviour.

TABLE G:3: NUMBER OF CIGARETTES CONDITIONAL MARGINAL EFFECTS FOR δ

	Model 1	**Model 2**	**Model 3**	**Model 4**
	Exponential	Hyperbolic	Quasi-Hyperbolic	Weibull
Number of cigarettes				
0	0.045 (0.012)	0.037 (0.010)	0.048 (0.013)	0.017 (0.005)
5	0.027 (0.007)	0.021 (0.006)	0.029 (0.008)	0.010 (0.003)
10	0.009 (0.005)	0.006 (0.004)	0.010 (0.005)	0.004 (0.002)
15	-0.010 (0.008)	-0.010 (0.008)	-0.009 (0.007)	-0.003 (0.002)
20	-0.028 (0.013)	-0.026 (0.013)	-0.028 (0.013)	-0.010 (0.004)
25	-0.046 (0.019)	-0.041 (0.018)	-0.047 (0.018)	-0.017 (0.006)

Standard errors in parentheses

In sum, the preceding results show that the smoking-discounting relationship is robust to the assumption that EU, rather than RDU, characterises choice under risk.

APPENDIX H
[ONLINE WORKING PAPER]

In this appendix we present the results from all of the two process mixture models that can be estimated from the four discounting functions used in the main text.[33] RDU and the Prelec PWF are assumed to characterise choice under risk in the joint estimation of these models. Focussing on a mixture of the E and H discounting models, we explain the statistical approach below and then present the results for all of the models.

Letting π^E represent the probability that the E model is correct, and $\pi^H = (1 - \pi^E)$ the probability that the H model is correct, the grand likelihood is the probability-weighted average of the two conditional likelihoods L^E and L^H for the E and H models, respectively. Thus, the likelihood for the mixture model is given by:

$$\ln L_i(r, \varphi, \eta, \delta_E, \delta_H, \mu, \nu, \kappa; z, X) = \Sigma_i \ln [(\pi^E \times L^E) + (\pi^H \times L^H)], \qquad (1)$$

where κ is a parameter which defines the log odds of the probability of the E model: $\pi^E = 1 / (1 + \exp(\kappa))$. This transformation allows the parameter κ to take on any value during the maximisation process but constrains the probabilities π^E and π^H to lie within the unit interval. The grand likelihood in (1) is maximised to estimate the parameters of each model and the weight accorded to each model in the data, under the assumptions that RDU and the Prelec PWF characterise choice under risk.

Table H:1 presents estimates of the mixture model of the E and H discounting functions assuming homogenous preferences. The estimate of $\pi^E = 0.347$ implies that the E model accounts for approximately 35% of the choices in the data; the H model therefore accounts for roughly 65% of the choices. A hypothesis test that $\pi^E = 0.5$ is

[33] One can use Vuong (1989) and Clarke (2007) non-nested model selection tests to formally determine which discounting function, in a pairwise comparison, finds more support in the dataset as a whole. The choice between these tests is based on the distribution of the individual log-ratios of the models. When the distribution of these log-ratios is leptokurtic, as we find in our data, the Clarke (2007) test is superior, from both statistical efficiency and power perspectives, to the Vuong (1989) test. Clarke (2007) tests of the four discounting models provided the following transitive ranking: WB > H > E > QH. In words, the WB model finds the most support in the data, the QH model finds the least support, and the E and H models are intermediate to these. Thus, if one had to select a single model that best characterises the time preferences of this sample, the WB model would be the choice. However, when multiple time preference processes are present in a dataset, one should be cognisant of this fact and estimate a mixture model to determine the proportion of choices best explained by each process. If one discounting model is truly superior to another, in the sense that it better explains all the data, then this will be reflected in a mixture probability estimate of zero for the inferior model.

easily rejected ($p < 0.001$) but so too is the hypothesis that $\pi^E = 0$ ($p < 0.001$).[34] Thus the E and H discounting models both find significant support in the data, even though the H model finds more support. Consequently it is a mistake to assume that only one DGP characterises the data.

TABLE H:1: MIXTURE MODEL ML ESTIMATES
RANK-DEPENDENT UTILITY AND HOMOGENOUS PREFERENCES

	Estimate	Std Error	p-value	95% Confidence Interval	
		Rank-dependent utility theory			
Power function parameter (r)	0.336***	0.027	0.000	0.283	0.390
PWF parameter (φ)	0.797***	0.025	0.000	0.749	0.846
PWF parameter (η)	0.893***	0.035	0.000	0.825	0.961
		Exponential discounting model			
Discounting parameter (δ_E^{mix})	0.137***	0.017	0.000	0.104	0.169
Mixture probability (π^E)	0.347***	0.034	0.000	0.280	0.414
		Hyperbolic discounting model			
Discounting parameter (δ_H^{mix})	0.730***	0.069	0.000	0.596	0.865
Mixture probability (π^H)	0.653***	0.034	0.000	0.586	0.720
		Error terms			
Risk Error (μ)	0.167***	0.008	0.000	0.152	0.182
Time Error (ν)	0.051***	0.015	0.001	0.021	0.081
N	17500				
log-likelihood	-8808.992				

H_0: $\pi^E = 0.5$, p-value < 0.001

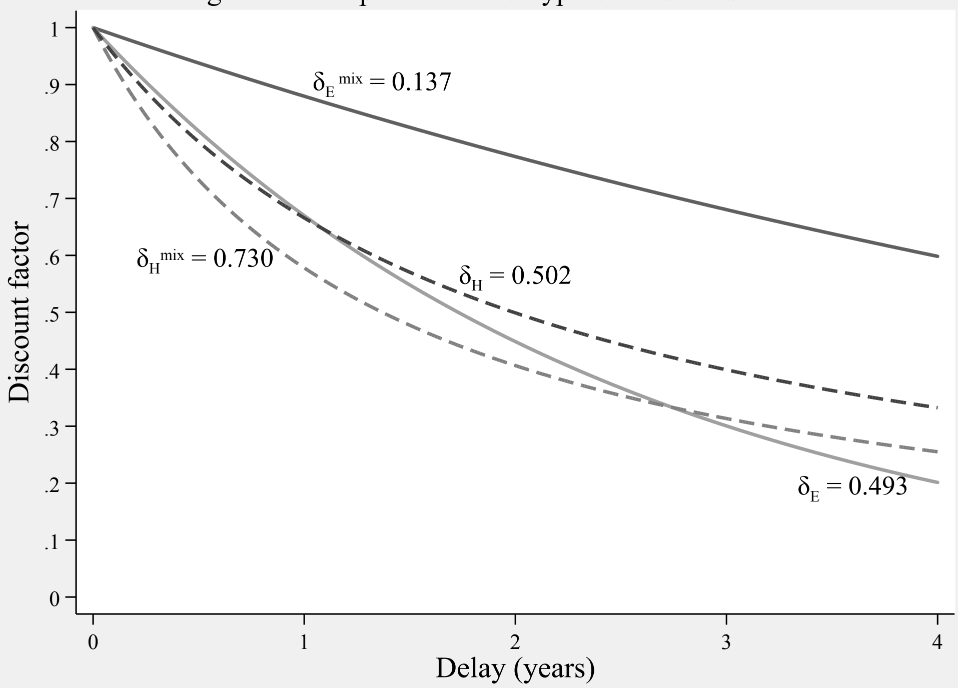
Results account for clustering at the individual level
* p<0.10, ** p<0.05, *** p<0.01

The mixture model in Table H:1 also shows how discounting parameter estimates are distorted when the E or H models have to account for all the data. In Model 1 of Table

[34] In all of the mixture models in this section, one of the discounting functions explains significantly more of the choices than the other discounting function. However, all discounting functions find significant support in the data which reinforces the point that it is a mistake to assume only one DGP characterises all discounting choices all of the time. The mixture probability estimates for the other mixture models are: E-QH model - $\pi^E = 0.636$; E-WB model - $\pi^E = 0.406$; H-QH model - $\pi^H = 0.634$; H-WB model - $\pi^H = 0.611$; QH-WB model - $\pi^{QH} = 0.609$.

F:2, where the E model was assumed to be the sole DGP, the estimate of $\delta_E = 0.493$. In the mixture model, the estimate of δ_E, which we refer to as δ_E^{mix} in Table H:1 and Figure H:1, is statistically significantly lower at 0.137 ($p < 0.001$). This implies that when one tries to make all the data fit the E model, one inflates the estimate of the discounting parameter since 65% of the data "wants" to be modelled as H. Similarly, in Model 2 of Table F:2, where the H model was assumed to be the sole DGP, the estimate of $\delta_H = 0.502$. In the mixture model in Table H:1, the estimate of δ_H^{mix} is statistically significantly higher at 0.730 ($p < 0.001$). Thus, by assuming one DGP we are averaging the estimates that we derive when allowing multiple DGPs to characterise the data.

Figure H:1: Exponential and hyperbolic discount factors



Finally, the estimate of the Fechner error term $\nu = 0.051$ in the mixture model in Table H:1 is statistically significantly lower than the estimates of ν for the E ($p < 0.001$) and H models ($p < 0.001$) in Table F:2. Thus, what was being captured as subject errors in decision making when estimating the E and H models separately is partly the product of forcing the data to fit one DGP.

We present the results from the other two process mixture models below. The take-home message is that all discounting models find significant support in the data.

TABLE H:2: MIXTURE MODEL ML ESTIMATES
RANK-DEPENDENT UTILITY AND HOMOGENOUS PREFERENCES

	Estimate	Std Error	*p*-value	95% Confidence Interval	
			Rank-dependent utility theory		
Power function parameter (r)	0.293***	0.027	0.000	0.240	0.346
PWF parameter (φ)	0.797***	0.025	0.000	0.748	0.846
PWF parameter (η)	0.853***	0.032	0.000	0.790	0.916
			Exponential discounting model		
Discounting parameter (δ_E^{mix})	0.815***	0.107	0.000	0.605	1.026
Mixture probability (π^E)	0.636***	0.039	0.000	0.561	0.712
			Quasi-Hyperbolic discounting model		
Discounting parameter (δ_{QH}^{mix})	0.105***	0.014	0.000	0.077	0.133
Discounting parameter (β_{QH}^{mix})	0.994***	0.002	0.000	0.991	0.998
Mixture probability (π^{QH})	0.364***	0.039	0.000	0.288	0.439
			Error terms		
Risk Error (μ)	0.175***	0.008	0.000	0.158	0.191
Time Error (ν)	0.034***	0.010	0.001	0.015	0.053
N	17500				
log-likelihood	-8775.861				

H_0: $\pi^E = 0.5$, *p*-value < 0.001

Results account for clustering at the individual level
* *p*<0.10, ** *p*<0.05, *** *p*<0.01

TABLE H:3: MIXTURE MODEL ML ESTIMATES
RANK-DEPENDENT UTILITY AND HOMOGENOUS PREFERENCES

	Estimate	Std Error	*p*-value	95% Confidence Interval	
		Rank-dependent utility theory			
Power function parameter (r)	0.301***	0.027	0.000	0.249	0.353
PWF parameter (φ)	0.797***	0.025	0.000	0.748	0.846
PWF parameter (η)	0.860***	0.032	0.000	0.798	0.923
		Exponential discounting model			
Discounting parameter (δ_E^{mix})	0.153***	0.017	0.000	0.119	0.186
Mixture probability (π^E)	0.406***	0.034	0.000	0.339	0.474
		Weibull discounting model			
Discounting parameter (δ_{WB}^{mix})	0.534***	0.131	0.000	0.279	0.790
Discounting parameter (β_{WB}^{mix})	5.940**	2.722	0.029	0.605	11.275
Mixture probability (π^{WB})	0.594***	0.034	0.000	0.526	0.661
		Error terms			
Risk Error (μ)	0.173***	0.008	0.000	0.157	0.189
Time Error (ν)	0.055***	0.014	0.000	0.027	0.084
N	17500				
log-likelihood	-8720.160				

H_0: $\pi^E = 0.5$, *p*-value = 0.006

Results account for clustering at the individual level
* *p*<0.10, ** *p*<0.05, *** *p*<0.01

TABLE H:4: MIXTURE MODEL ML ESTIMATES
RANK-DEPENDENT UTILITY AND HOMOGENOUS PREFERENCES

	Estimate	Std Error	*p*-value	95% Confidence Interval	
		Rank-dependent utility theory			
Power function parameter (r)	0.329***	0.027	0.000	0.277	0.381
PWF parameter (φ)	0.797***	0.025	0.000	0.749	0.846
PWF parameter (η)	0.886***	0.035	0.000	0.819	0.954
		Hyperbolic discounting model			
Discounting parameter (δ_H^{mix})	0.717***	0.067	0.000	0.585	0.849
Mixture probability (π^H)	0.634***	0.036	0.000	0.564	0.704
		Quasi-Hyperbolic discounting model			
Discounting parameter (δ_{QH}^{mix})	0.119***	0.015	0.000	0.090	0.148
Discounting parameter (β_{QH}^{mix})	0.994***	0.002	0.000	0.990	0.998
Mixture probability (π^{QH})	0.366***	0.036	0.000	0.296	0.436
		Error terms			
Risk Error (μ)	0.168***	0.008	0.000	0.153	0.184
Time Error (ν)	0.046***	0.012	0.000	0.021	0.070
N	17500				
log-likelihood	-8752.007				

H_0: $\pi^H = 0.5$, *p*-value < 0.001

Results account for clustering at the individual level
* *p*<0.10, ** *p*<0.05, *** *p*<0.01

TABLE H:5: MIXTURE MODEL ML ESTIMATES
RANK-DEPENDENT UTILITY AND HOMOGENOUS PREFERENCES

	Estimate	Std Error	*p*-value	95% Confidence Interval	
			Rank-dependent utility theory		
Power function parameter (r)	0.328***	0.026	0.000	0.276	0.379
PWF parameter (φ)	0.797***	0.025	0.000	0.749	0.846
PWF parameter (η)	0.884***	0.034	0.000	0.817	0.952
			Hyperbolic discounting model		
Discounting parameter (δ_H^{mix})	0.720***	0.069	0.000	0.585	0.855
Mixture probability (π^H)	0.611***	0.039	0.000	0.535	0.688
			Weibull discounting model		
Discounting parameter (δ_{WB}^{mix})	0.072***	0.009	0.000	0.053	0.090
Discounting parameter (β_{WB}^{mix})	1.759***	0.206	0.000	1.355	2.164
Mixture probability (π^{WB})	0.389***	0.039	0.000	0.312	0.465
			Error terms		
Risk Error (μ)	0.169***	0.008	0.000	0.153	0.184
Time Error (ν)	0.044***	0.012	0.000	0.021	0.068
N	17500				
log-likelihood	-8703.874				

H_0: $\pi^H = 0.5$, *p*-value = 0.004

Results account for clustering at the individual level
* $p<0.10$, ** $p<0.05$, *** $p<0.01$

TABLE H:6: MIXTURE MODEL ML ESTIMATES
RANK-DEPENDENT UTILITY AND HOMOGENOUS PREFERENCES

	Estimate	Std Error	*p*-value	95% Confidence Interval	
		Rank-dependent utility theory			
Power function parameter (r)	0.291***	0.027	0.000	0.238	0.343
PWF parameter (φ)	0.797***	0.025	0.000	0.748	0.846
PWF parameter (η)	0.851***	0.032	0.000	0.788	0.913
		Quasi-Hyperbolic discounting model			
Discounting parameter (δ_{QH}^{mix})	0.797***	0.111	0.000	0.579	1.015
Discounting parameter (β_{QH}^{mix})	0.996***	0.003	0.000	0.990	1.002
Mixture probability (π^{QH})	0.609***	0.044	0.000	0.524	0.695
		Weibull discounting model			
Discounting parameter (δ_{WB}^{mix})	0.066***	0.009	0.000	0.047	0.084
Discounting parameter (β_{WB}^{mix})	1.720***	0.188	0.000	1.352	2.087
Mixture probability (π^{WB})	0.391***	0.044	0.000	0.305	0.476
		Error terms			
Risk Error (μ)	0.175***	0.008	0.000	0.159	0.192
Time Error (ν)	0.034***	0.010	0.000	0.015	0.052
N	17500				
log-likelihood	-8715.090				

H_0: $\pi^{QH} = 0.5$, *p*-value = 0.012

Results account for clustering at the individual level
* *p*<0.10, ** *p*<0.05, *** *p*<0.01

ADDITIONAL REFERENCES

CLARKE, K. A. (2007): “A Simple Distribution-Free Test for Nonnested Model Selection,” *Political Analysis*, 15, 347-363.

VUONG, Q. H. (1989): “Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses,” *Econometrica*, 57, 307-333.